Identifying Genetic Variants and Environmental Factors Influencing Multiple Sclerosis Pathological Processes

By

Yuan Zhou

B.A., Biology (CHINA);
M.S., Genetics (CHINA)

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

University of Tasmania, Hobart

September 2016

Declaration of originality

This thesis contains no material which has been accepted for a degree or diploma by the University of Tasmania, nor any other institution, except by way of background information and duly acknowledged in the thesis, and to the best of my knowledge and belief, no material previously published or written by another person except where due acknowledgement is made in the text of the thesis

Signature

22 Sep 2016

Yuan Zhou Date

Statement of authority to access

This thesis may be available for loan. Copying of any part of this thesis is prohibited for two years from the date this statement was signed; after that time, limited copying is permitted in accordance with the *Australian Copyright Act of 1968.*

Signature

22 Sep 2016

Yuan Zhou. Date

Statement of co-authorship

This thesis includes papers for which Yuan Zhou (YZ) is not sole author. YZ took the lead in this research, developing and implementing the analyses included herein under the supervision of Bruce V Taylor (BVT), Jac Charlesworth (JC) and Steve Simpson, Jr. (SSJ). In this process, however, he was assisted by co-authors to varying extent. Following then, the contributions of each co-author are detailed for each respective project.

The paper reported in Chapter 1:

Zhou Y, Simpson S, Jr., Holloway AF, Charlesworth J, van der Mei I, Taylor BV. The potential role of epigenetic modifications in the heritability of multiple sclerosis. Mult Scler. 2014;20(2):135-40.

YZ undertook the literature review with direct assistance from all authors. All authors contributed equally to the development of the review.

The paper reported in Chapter 2:

Zhou Y, Zhu G, Charlesworth J, Simoson S, Jr., Rubicz R, Göring H, Patsopoulos AN, Laverty C, Wu FT, Henders A, Ellis J, van ver Mei I, Montogomery G, Blangero J, Curran J, Johoson M, Martin N, Nyholt D and Taylor BV. Genetic loci for Epstein-Barr Virus nuclear antigen-1 are associated with risks of multiple sclerosis. Mult Scler.2016

In concert with and under guidance of BVT, Nicholas G. Martin (NGM), Dale R. Nyholt (DRN), JC and SSJ, YZ developed and implemented all statistical analyses. YZ composed the drafts of the manuscript and coordinated revision.

GZ implemented the statistical analyses in QIMR Twin family GWAS, and was involved in the initial drafting and critical revision of the manuscript.

JC provided guidance with statistical analyses undertaken in this study, and was involved in critical revision of the manuscript.

SSJ provided guidance with statistical analyses undertaken in this study, and was involved in the initial drafting and critical revision of the manuscript.

Rohina Rubicz (RR) provided the summary results of the Mexican Amercian families EBNA-1 GWAS, and was involved in critical revision of the manuscript.

Harald H.H.Göring (HHHG) was involved in critical revision of the manuscript.

Nikolaos A. Patsopoulos (NAP) provided the summary results of the MS GWAS meta analysis results, and was involved in critical revision of the manuscript.

Caroline Laverty (CL) did the lab work by measuring the anti-EBNA-1 IgG titers in QIMR Twin families.

Feitong Wu (FW) provided help with statistical analyses undertaken in this study, and was involved in critical revision of the manuscript.

Anjali Henders (AH) provided help by rechecking the results of anti-EBNA-1 IgG titers in QIMR Twin families, and was involved in critical revision of the manuscript.

Jonathan J.Ellis (JJE) provided the guidance regarding to the cis-eQTL analysis, and was involved in critical revision of the manuscript.

Ingrid van der Mei (IvM) contributed to the critical revision of the manuscript.

Grant W.Montgomery (GWM) contributed to the critical revision of the manuscript.

John Blangero provided the guidance with statistical analyses, and was involved in critical revision of the manuscript.

Joanne E.Curran contributed to the critical revision of the manuscript.

Matthew P. Johnson contributed to the critical revision of the manuscript.

ANZgene consortium provided access to the data relating to the gene-gene interaction analysis and was involved in the critical revision of the manuscript.

NGM provided the guidance with statistical analyses, and was involved in critical revision of the manuscript.

DRN provided the guidance with statistical analyses, and was involved in critical revision of the manuscript.

BVT worked in concert with NGM, DRN to conceive and design this study, develop analyses undertaken in this study, and is involved in the initial drafting and critical revision of the manuscript.

The paper reported in Chapter 3:

Zhou Y, Taylor B, van der Mei I, Stewart N, Charlesworth J, Blizzard L, Ponsonby AL, Dwyer T, Pittas F and Simpson SJ. Genetic variation in PBMC-produced IFN-gamma and TNF-alpha associations with relapse in multiple sclerosis. J Neurol Sci. 2015 Feb 15;349(1-2):40-4.

YZ was involved in the development and implementation of statistical analyses undertaken, under supervision by SSJ, IvM, JC and BVT. YZ composed drafts of the manuscript and coordinated revision.

BVT was involved in the development and acquisition of funding for the MS Longitudinal Study from which the data for this analysis was drawn, along with Anne- Louise Ponsonby (A-LP), Fotini Pittas (FP), Terence Dwyer (TD) and IvM. BVT was involved in the critical revision of the manuscript.

IvM was involved in the development and acquisition of funding for the MS Longitudinal Study, along with BVT, A-LP, FP, and TD. IvM was involved in the data collection for the MS Longitudinal Study along with BT and FP. IvM was involved in conception of the analyses used. IvM was involved in the critical revision of the manuscript.

Niall Stewart (NS) contributed to the critical revision of the manuscript.

JC provided guidance and supervision for the genetic statistical analyses undertook in this study, and was involved in the critical revision of the manuscript.

Leigh Blizzard (LB) provided guidance and supervision for the statistical analyses undertook in this study, and was involved in the critical revision of the manuscript.

A-LP was involved in the development and acquisition of funding for the MS Longitudinal Study, along with BVT, FP, TD, and IvM. A-LP contributed to the critical revision of the manuscript.

TD was involved in the development and acquisition of funding for the MS Longitudinal Study, along with BVT, A-LP, FP, and IvM. TD contributed to the critical revision of the manuscript.

FP was involved in the development and acquisition of funding for the MS Longitudinal Study, along with BVT, A-LP, TD, and IvM. FP was involved in the data collection for the MS Longitudinal Study along with BT and IvM. FP contributed to the critical revision of the manuscript.

SSJ conceived and designed this study, developed analyses undertaken in this study, and was involved in the initial drafting and critical revision of the manuscript.

The paper reported in Chapter 4:

Zhou Y, Simpson SJ, Charlesworth J, van der Mei I, Lucas R, Ponsonby AL,AUSLONG investigators group, Taylor B. Variation within the *MBP* gene predicts disease course in multiple sclerosis.

YZ was involved in the development and implementation of statistical analyses undertaken, under supervision by SSJ, JC and BVT. YZ composed drafts of the manuscript and coordinated revision.

SSJ provided guidance with statistical analyses undertaken in this study, and was involved in the initial drafting and critical revision of the manuscript.

IvM was involved in the development and acquisition of funding for the Auslong Study, along with A-LP, Robyn Locus (RL) and BVT. IvM was involved in the data collection for Auslong Study along with A-LP, RL and BVT. IvM was involved in conception of the analyses used. IvM was involved in the critical revision of the manuscript.

RL was involved in the development and acquisition of funding for the Auslong study, along with IvM, A-LP and BVT. RL was involved in the data collection for the Auslong Study along with IvM, A-LP and BVT. RL was involved in the critical revision of the manuscript.

A-LP was involved in the development and acquisition of funding for the Auslong study, along with IvM, RL and BVT. A-LP was involved in the data collection for the Auslong Study along with IvM, RL and BVT. A-LP was involved in the critical revision of the manuscript.

AUSLONG investigators group was involved in the development and acquisition of funding for the Auslong study, all the members were involved in the revision of the manuscript.

BVT was involved in the development and acquisition of funding for the Auslong study, along with IvM, RL and A-LP. BVT was involved in the data collection for the Auslong Study along with IvM, RL and A-LP. BVT conceived and designed this study, developed analyses undertaken in this study, and was involved in the initial drafting and critical revision of the manuscript.

The paper reported in Chapter 5

Zhou Y, Chen M, Simpson SJ, Charlesworth J, van der Mei I, Lucas R, Ponsonby AL,AUSLONG investigators group, Taylor B. Common Genetic Variation within the *miR-146a* predicts disease onset & relapse in multiple sclerosis.

YZ was involved in the development and implementation of statistical analyses undertaken, under supervision by SSJ, JC and BVT. YZ composed drafts of the manuscript and coordinated revision.

Ming Chen (MC) was involved in the development and implementation of statistical analyses, and was involved in the initial drafting and critical revision of the manuscript.

SSJ provided guidance with statistical analyses undertaken in this study, and was involved in the initial drafting and critical revision of the manuscript.

IvM was involved in the development and acquisition of funding for the Auslong Study, along with A-LP, Robyn Locus (RL) and BVT. IvM was involved in the data collection for Auslong Study along with A-LP, RL and BVT. IvM was involved in

conception of the analyses used. IvM was involved in the critical revision of the manuscript.

RL was involved in the development and acquisition of funding for the Auslong study, along with IvM, A-LP and BVT. RL was involved in the data collection for the Auslong Study along with IvM, A-LP and BVT. RL was involved in the critical revision of the manuscript.

A-LP was involved in the development and acquisition of funding for the Auslong study, along with IvM, RL and BVT. A-LP was involved in the data collection for the Auslong Study along with IvM, RL and BVT. A-LP was involved in the critical revision of the manuscript.

AUSLONG investigators group was involved in the development and acquisition of funding for the Auslong study, all the members were involved in the revision of the manuscript.

BVT was involved in the development and acquisition of funding for the Auslong study, along with IvM, RL and A-LP. BVT was involved in the data collection for the Auslong Study along with IvM, RL and A-LP. BVT conceived and designed this study, developed analyses undertaken in this study, and was involved in the initial drafting and critical revision of the manuscript.

Statement of co-authorship

Signed

_____

Yuan Zhou

Date 22 Sep 2016

Statement of ethical conduct

The research associated with this thesis abides by the international and Australian codes on human and animal experimentation, the guidelines by the Australian Government's Office of the Gene Technology Regulator and the rulings of the Safety, Ethics and Institutional Biosafety Committees of the University.

Signed

Yuan Zhou

Date 22 Sep 2016

Acknowledgement

I am greatly appreciative for the thorough and patient guidance and support I received from my supervisors, Bruce, Jac, Ingrid & Steve, and also collaborators from QIMR Berghofer like Gu, Nick and Dale, throughout the duration of my PhD. Without them, I couldn't have completed my projects, not to mention to develop my own ideas and to grow as an independent researcher.

Being a doctoral student is always stressful in work. Fortunately, I have a warm family that provide constant love, concern and full support. I owe them a lot and will always be grateful for how lucky I am to have them.

Lastly, I would like to acknowledge the participants of the QIMR twin families, MS Longitudinal Study and the Ausimmune/AusLong Longitudinal Study, as well as the numerous participants in the studies included in the meta-analysis. It is their kind participation that makes all the research work possible, so I hope that the achievements of my work will be of benefit for them and all MS patients.

Abstract

Multiple sclerosis (MS) is a complex neurological disorder whose cause and subsequent clinical course are to a large part due to an interplay between environmental and genetic factors. Currently, apart from the well-known effects of MHC genetic risk loci, genome-wide association studies (GWAS) have identified 110 non-MHC risk loci. However, these findings can only explain part of the heritability of MS, leading to the question "where is the missing heritability". This thesis presents a number of methodological analyses using different study designs to reveal, in part, the puzzle of the missing heritability. By examining the susceptibility genes for EBV infection and their roles in MS development, as well as evaluating genetic variations and key environmental factors in determining clinical course, the complex role of genetics in MS onset and progression might be better understood.

In Chapter 2, I performed a genome-wide association study (GWAS) of anti-Epstein Barr Nuclear Antigen-1 (EBNA-1) IgG titers in 3,599 individuals from an unselected twin family cohort, followed by a meta-analysis with data from an independent EBNA-1 GWAS. I then examined the shared polygenic risk between the EBNA-1 GWAS (effective sample size (N_{eff}) = 5,555) and a large MS GWAS (N_{eff} = 15,231). I identified one locus of strong association within the human leukocyte antigen (HLA) region that associated with anti-EBNA-1 titers. By examining the shared polygenic risk, I show that apart from the confirmed HLA region, the association of anti-EBNA-1 IgG titer with MS risk is also mediated through non-HLA loci (1p22.1, 3p24.1, 3q13.33 and 10p15.1).

In Chapter 3, I assessed whether SNPs within genes for relevant cytokines and their receptors modulate the associations of TNF-α and IFN-γ with relapse. I found variation within the gene, *IL2RB,* demonstrated a significant protective effect of TNF-α on relapse, while

variation in the 3' region of *TNFRSF1B* and variation in the 5'region of *IL3* showed a strong association between IFN-γ and increased relapse risk.

In Chapter 4, I assessed whether genetic variation in the myelin basic protein gene may directly, or by interaction with anti-HHV6 or EBV serological titers, determine clinical outcomes (conversion to MS after a first demyelinating event (FDE), relapse rate and disability) using the Ausimmune/AusLong Study. I found one variant, rs12959006, predicted worse MS clinical course in three key metrics, namely conversion to MS, relapse risk and annualised disability progression. I also found a significant interaction between the risk genotype and baseline anti-HHV6 IgG in predicting conversion to MS and relapse. These results, if replicated, may aid in developing prognostic algorithms in the early disease period in MS as well as providing further mechanistic insights.

In Chapter 5, I addressed the key question regarding the role of epigenetics in the missing heritability of MS. For the first time, I assessed the effects of a functional genetic variant within the gene for miR-146a (a miRNA whose expression is strongly associated with MS disease activity) on MS clinical course in the Ausimmune/AusLong Study, a well-established FDE cohort. I found the risk genotype (GC+CC) of the miR-146a SNP rs2910164 not only itself significantly predicted relapse risk, but also showed significant additive interaction with markers of EBV infection in predicting CDMS and relapse, including anti-EBNA-1 and -2 IgG and history of infectious mononucleosis, such that the associations of these environmental risk factors were significantly more potent among those carrying the miR-146a risk variant.

This thesis presents a range of studies that add significant information regarding the missing heritability of MS and provide more evidence for the role of classic genetic, epigenetic and environmental factors in the pathogenesis of MS. Additionally, this thesis demonstrates how

genetic variants may influence MS clinical course and interact with environmental factors to enhance the effects of genetic variants in predicting critical MS progression metrics. This work will be useful for the scientific community in general, and will be of benefit for MS patients.

Table of Contents

List of Tables

List of Supplemental Tables

List of Figures

Lists of Supplementary Figures

Papers directly arising from the work described in this thesis

Papers published

Chapter 1:

Zhou Y, Simpson S, Jr., Holloway AF, Charlesworth J, van der Mei I, Taylor BV. The potential role of epigenetic modifications in the heritability of multiple sclerosis. Mult Scler. 2014;20(2):135-40.

Chapter 2:

Zhou Y, Zhu G, Charlesworth J, Simoson S, Jr., Rubicz R, Göring H, Patsopoulos AN, Laverty C, Wu FT, Henders A, Ellis J, van ver Mei I, Montogomery G, Blangero J, Curran J, Johoson M, Martin N, Nyholt D and Taylor BV. Genetic loci for Epstein-Barr Virus nuclear antigen-1 are associated with risks of multiple sclerosis. Mult Scler.2016

Chapter 3:

Zhou Y, Taylor B, van der Mei I, Stewart N, Charlesworth J, Blizzard L, Ponsonby AL, Dwyer T, Pittas F and Simpson SJ. Genetic variation in PBMC-produced IFN-gamma and TNF-alpha associations with relapse in multiple sclerosis. Journal of the neurological sciences. 2015 Feb 15;349(1-2):40-4.

Conference presentations arising from work in this thesis

- Two oral presentations in MSRA meeting at Melbourne, 28-30 Oct 2015
- Two poster presentations in ECTRIMS meeting at Barcelona, 07-10 Oct 2015
- One oral and one poster presentations at PACTRIMS meeting, Taipei, 06-08 Nov 2014
- Two poster presentations in 2014 Joint ACTRIMS-ECTRIMS Meeting Boston, USA, 10-13 Sep 2014
- One oral presentation at GeneMappers meeting 2014 at Adelaide, 11-14 May 2014
- One oral presentation as finalist for ASMR Medical Research Week, 2014 Tasmanian Postgraduate Student Research Awards

Awards received from the work described in this thesis

- Educational Grant Award provided by ECTRIMS meeting at Barcelona, 07-10 Oct 2015
- Travel Grant Award provided by China CDC for PACTRIMS meeting, Taipei, 06-08 Nov 2014
- Educational Grant Award provided by 2014 Joint ACTRIMS-ECTRIMS Meeting at Boston, USA, 10-13 Sep 2014

List of abbreviations

Abbreviation	Full term
µg	Micrograms
$1,25(OH)_2D$	1,25-dihydroxyvitamin D
25(OH)D	25-hydroxyvitamin D
95% CI	95 percent Confidence Interval
APC	Antigen Presenting Cell
BMI	Body Mass Index
CDMS	Clinical definite Multiple Sclerosis
CNS	Central Nervous System
CSF	Cerebrospinal fluid
DIS	Dissemination in space
DIT	Dissemination in time
EBNA	Epstein-Barr Nuclear Antigen
EBV	Epstein-Barr Virus
EDSS	Kurtzke Expanded Disability Severity Scale
ELISA	Enzyme-linked Immunosorbent Assay
FDE	First demyelination event
GRS	Genetic risk score
GWAS	Genome wide association analysis
HERV	Human Endogenous Retrovirus
HHV	Human Herpesvirus
HHV-6	Human Herpesvirus 6
HLA	Human Leukocyte Antigen
HLA-DRB1	Major Histocompatibility Complex, Class II, DR beta 1
HR	Hazard Ratio
IFN-β	Interferon beta
IgG	Immunoglobulin class G
IgM	Immunoglobulin class M
IL	Interleukin
IQR	Interquartile Range
IU	International Units
Km	Kilometers
LD	Linkage disequilibrium
MBP	Myelin basic protein
MRI	Magnetic Resonance Imaging
MS	Multiple Sclerosis
MSL	Multiple Sclerosis Longitudinal Study
MSSS	Multiple Sclerosis Severity Score
nmol/L	Nanomoles per Liter
PBMC	Peripheral Blood Mononuclear Cells
PCR	Polymerase Chain Reaction
PPMS	Primary-Progressive Multiple Sclerosis
RCT	Randomised Controlled Trial
RRMS	Relapsing-Remitting Multiple Sclerosis
SD	Standard Deviation
SPMS	Secondary-Progressive Multiple Sclerosis

List of abbreviations

SNP	Single nucleotide polymorphism
T_h1	Helper T-lymphocyte class 1
T_h2	Helper T-lymphocyte class 2
T_h17	Helper T-lymphocyte class 17
T_{reg}	Regulatory T-lymphocyte
UK	United Kingdom of Great Britain and Northern Ireland
USA	United States of America
UVR	Ultraviolet radiation
VCA	Viral Capsid Antigen

Chapter 1: Multiple sclerosis background and literature review

Multiple sclerosis (MS) is a chronic inflammatory autoimmune disease of the central nervous system (CNS) with pathological hallmarks of inflammation, demyelination, remyelination, and neurodegeneration. Although not fully delineated, its cause is thought to be an interplay between environmental and genetic factors[1,2].

MS was identified as an important clinical disorder by Jean-Martin Charcot in 1868[3], when he described three patients who had the symptoms subsequently associated with the disease dating back to 1855; he effectively gathered the first clinical and pathological observations of the condition. A landmark paper by James Dawson in the early 20th century marked the beginning of an improved understanding of MS etiology and pathogenesis[4]; he described the disease as "classically exhibited in irregular, reversible attacks, while the proportion of axonal loss could be observed in autopsy tissue sections as resulting from an inflammatory action of microglia". Another important discovery in the comprehension of MS was the discovery in the 1930s by River and colleagues of the animal brain inflammation model experimental autoimmune encephalomyelitis (EAE)[5]. This model, although not a perfect analogue to human pathogenesis, nonetheless ushered in an era of the discovery of the molecular mechanisms underlying the inflammatory process involved in human CNS demyelinating diseases. A number of terms for the condition have been used, including disseminated sclerosis and encephalomyelitis disseminate, but in the 1950s, a consistency of nomenclature for MS was finally achieved with the publication of *Multiple Sclerosis* by Douglas McAlpine, Nigel Compston, and Charles Lumsden[6].

In the 21st century, various technologies have allowed further understanding of MS, both in terms of diagnostic mechanisms in clinical practice and pathogenesis. For example, the use of magnetic resonance imaging (MRI) has revolutionized the diagnosis of MS and is useful in evaluating neuroprotection in MS.

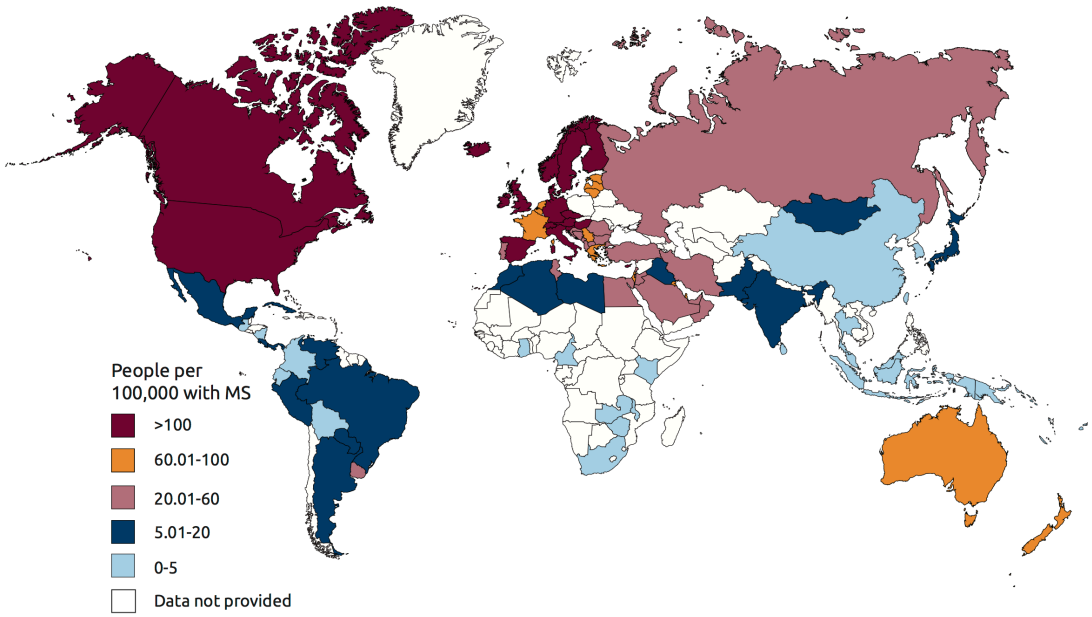
For identifying the role of genetic factors in the etiology of MS, perhaps the most powerful methods are genome-wide association studies. These studies typically evaluate millions of single nucleotide polymorphisms (SNPs) to identify MS susceptibility loci that modulate disease risk. However, currently discovered risk loci from such genome-wide association studies can explain only approximately 28% of the perceived heritability of MS, leading to the following question: where is the missing heritability? Moreover, in addition to the role of genetic factors in disease onset, the role of genetic determinants in clinical course, both in early and established MS, has not been as well studied. In this thesis, I present my efforts to address these aspects. This introductory chapter provides some background information for the understanding of the remainder of the thesis.

1.1 The epidemiology of MS

According to the latest report of the Atlas of MS 2013,[7] which assessed the epidemiology of MS around the world, the estimated number of people with MS was approximately 2.3 million, increasing by approximately 200,000 since 2008. MS is more than twice as common among women as men. Figure 1.1 shows the prevalence by country; countries with large populations of European descent have the highest prevalences. The average age of MS onset is 30 years; only 2-5% of people with MS

are diagnosed under age 18, and there is some suggestion that this pediatric-onset MS is a separate condition.

Figure 1.1: The prevalence of MS by country (reproduced from report[7]).



1.2 Clinical manifestation of MS

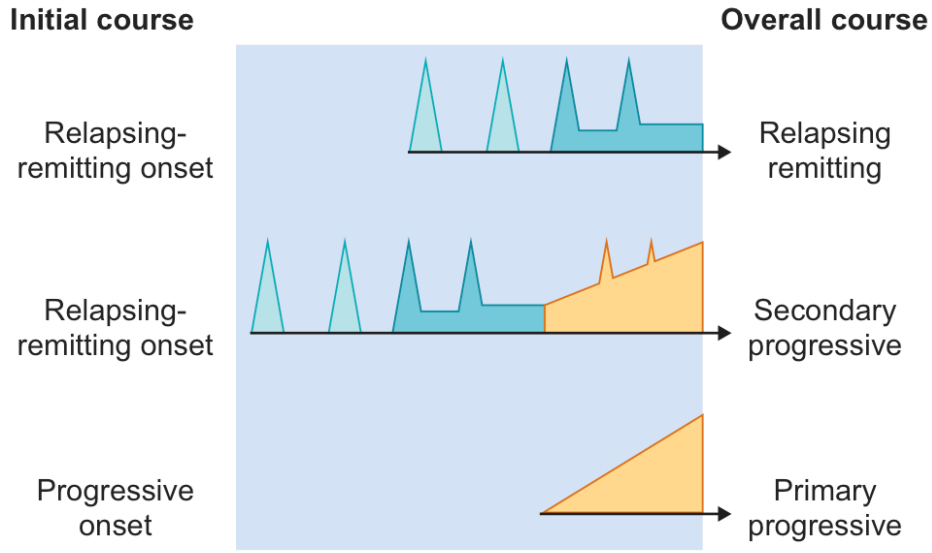
The course of MS can be characterized in terms of two distinctive clinical aspects: relapse and progression. Relapses, also known as exacerbations, attacks, or episodes, occur as a result of new focal inflammatory lesions or the reactivation of old lesions. Relapses are defined as the neurological symptom's first occurrence, recurrence, or worsening with clearly noticeable onset that lasts for a stable period followed by partial or complete recovery[8]. The clinical definition of relapse typically requires the event to last more than 24 hours in the absence of fever or infection[8].

Progression is defined as a persistent deterioration in neurologic function, which is particularly evident in the progressive phases of the disease, though it is also present

to an extent in the relapsing-remitting form[8]. The primary metric for assessing disability is the Kurtzke Expanded Disability Status Scale (EDSS), which assigns scores for the functioning of eight functional systems[9], and the combined score ranges from 0 (normal) to 10 (death due to MS) in increments of 0.5 units.

The classifications of MS clinical subtypes are based on the presence or absence of relapses and progression. The most common type is relapsing-remitting MS (RRMS), which is characterized by intermittent relapses during periods of relative clinical inactivity (remission). Approximately 65% of patients with RRMS convert to secondary progressive MS (SPMS), which features a fairly uniform progression rate without relapses. The least common subtype is primary progressive MS (PPMS), which manifests as a slow accumulation of disability from disease onset[10,11] (Figure 1.2).

Figure 1.2: Classification of the course of multiple sclerosis (reproduced from Compston and colleagues[8])



1.3 Diagnosis of multiple sclerosis

The diagnosis of MS is now based on the McDonald/Polman criteria, which include clinical and paraclinical assessments that are used to define requirements for demonstrating disease dissemination in space and dissemination in time. These criteria use a combination of clinical and paraclinical evidence to substantiate two demyelinating events separated in time and space within the central nervous system and for which there are no better explanations.

1.3.1 McDonald Criteria

The most commonly used diagnostic criteria are the McDonald/Polman Criteria, now in the 2010 revision iteration. For relapsing-remitting MS, this involves the demonstration of the dissemination of neurological demyelination events in space and time, whereas for the progressive-onset types, a sustained disability progression over

at least 12 months and supported by paraclinical evidence is necessary. Both of these diagnoses presume the exclusion of other potential explanatory mechanisms. The first iteration, the archetypal McDonald criteria, was presented in 2001 by the International Panel on the Diagnosis of Multiple Sclerosis[12] led by W. Ian McDonald. Subsequently, these criteria underwent two major revisions in 2005[13] and 2010[14] that served to simplify the diagnostic criteria, enhance the diagnostic sensitivity and specificity, make better use of improved MRI technology, and improve the applicability in other non-Caucasian populations (Figure 1.3).

Figure 1.3: 2010 revision to the McDonald criteria (reproduced from Polman and colleagues[14])

TABLE 4: The 2010 McDonald Criteria for Diagnosis of MS

Clinical Presentation	Additional Data Needed for MS Diagnosis
≥2 attacks[a]; objective clinical evidence of ≥2 lesions or objective clinical evidence of 1 lesion with reasonable historical evidence of a prior attack[b]	None[c]
≥2 attacks[a]; objective clinical evidence of 1 lesion	Dissemination in space, demonstrated by: ≥1 T2 lesion in at least 2 of 4 MS-typical regions of the CNS (periventricular, juxtacortical, infratentorial, or spinal cord)[d]; or Await a further clinical attack[a] implicating a different CNS site
1 attack[a]; objective clinical evidence of ≥2 lesions	Dissemination in time, demonstrated by: Simultaneous presence of asymptomatic gadolinium-enhancing and nonenhancing lesions at any time; or A new T2 and/or gadolinium-enhancing lesion(s) on follow-up MRI, irrespective of its timing with reference to a baseline scan; or Await a second clinical attack[a]
1 attack[a]; objective clinical evidence of 1 lesion (clinically isolated syndrome)	Dissemination in space and time, demonstrated by: For DIS: ≥1 T2 lesion in at least 2 of 4 MS-typical regions of the CNS (periventricular, juxtacortical, infratentorial, or spinal cord)[d]; or Await a second clinical attack[a] implicating a different CNS site; and For DIT: Simultaneous presence of asymptomatic gadolinium-enhancing and nonenhancing lesions at any time; or A new T2 and/or gadolinium-enhancing lesion(s) on follow-up MRI, irrespective of its timing with reference to a baseline scan; or Await a second clinical attack[a]
Insidious neurological progression suggestive of MS (PPMS)	1 year of disease progression (retrospectively or prospectively determined) plus 2 of 3 of the following criteria[d]: 1. Evidence for DIS in the brain based on ≥1 T2 lesions in the MS-characteristic (periventricular, juxtacortical, or infratentorial) regions 2. Evidence for DIS in the spinal cord based on ≥2 T2 lesions in the cord 3. Positive CSF (isoelectric focusing evidence of oligoclonal bands and/or elevated IgG index)

If the Criteria are fulfilled and there is no better explanation for the clinical presentation, the diagnosis is "MS"; if suspicious, but the Criteria are not completely met, the diagnosis is "possible MS"; if another diagnosis arises during the evaluation that better explains the clinical presentation, then the diagnosis is "not MS."

[a]An attack (relapse; exacerbation) is defined as patient-reported or objectively observed events typical of an acute inflammatory demyelinating event in the CNS, current or historical, with duration of at least 24 hours, in the absence of fever or infection. It should be documented by contemporaneous neurological examination, but some historical events with symptoms and evolution characteristic for MS, but for which no objective neurological findings are documented, can provide reasonable evidence of a prior demyelinating event. Reports of paroxysmal symptoms (historical or current) should, however, consist of multiple episodes occurring over not less than 24 hours. Before a definite diagnosis of MS can be made, at least 1 attack must be corroborated by findings on neurological examination, visual evoked potential response in patients reporting prior visual disturbance, or MRI consistent with demyelination in the area of the CNS implicated in the historical report of neurological symptoms.

[b]Clinical diagnosis based on objective clinical findings for 2 attacks is most secure. Reasonable historical evidence for 1 past attack, in the absence of documented objective neurological findings, can include historical events with symptoms and evolution characteristics for a prior inflammatory demyelinating event; at least 1 attack, however, must be supported by objective findings.

[c]No additional tests are required. However, it is desirable that any diagnosis of MS be made with access to imaging based on these Criteria. If imaging or other tests (for instance, CSF) are undertaken and are negative, extreme caution needs to be taken before making a diagnosis of MS, and alternative diagnoses must be considered. There must be no better explanation for the clinical presentation, and objective evidence must be present to support a diagnosis of MS.

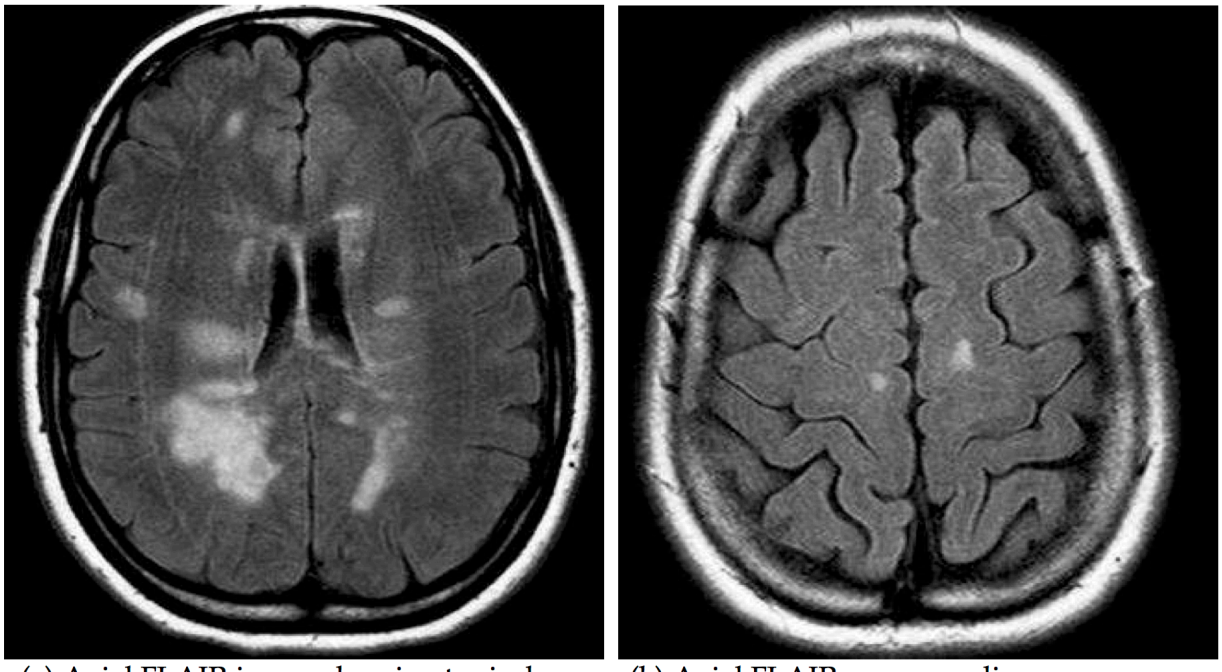
[d]Gadolinium-enhancing lesions are not required; symptomatic lesions are excluded from consideration in subjects with brainstem or spinal cord syndromes.

MS = multiple sclerosis; CNS = central nervous system; MRI = magnetic resonance imaging; DIS = dissemination in space; DIT = dissemination in time; PPMS = primary progressive multiple sclerosis; CSF = cerebrospinal fluid; IgG = immunoglobulin G.

1.3.2 The use of conventional MRI in the diagnosis of MS

Conventional MRI, which uses magnetic fields and radio waves to measure brain lesion number, volume, and changes over time, has become a critical tool in the diagnosis of MS. The widely used techniques for conventional MRI include T2-weighted, fast fluid-attenuated inversion recovery (FLAIR) and T1-weighted imaging, with and without contrast (gadolinium) administration[15]. Compared with T2-weighted MRI, FLAIR imaging has greater sensitivity for detecting white matter hyperintensities in the periventricular and cortical/juxtacortical regions, which are shown as bright areas against a gray background[15] (Figure 1.4).

Figure 1.4: Axial and sagittal FLAIR images showing common sites of MS involvement (reproduced from Andreadou[15]).



(a) Axial FLAIR image showing typical ovoid periventricular hyperintense MS lesions

(b) Axial FLAIR scan revealing juxtacortical lesions

1.4 The immunopathogenesis of MS

The immune system plays an important role in the development of MS. The human leukocyte antigen (HLA) locus, located on chromosome 6, is the most well-established MS genetic risk factor; the HLA class II allele DRB1*1501 has the strongest association of any genetic determinant studied (OR=3.1) for the development of MS[16]. The HLA genes encode the major histocompatibility complex (MHC) proteins MHC I and II, which are essential components of antigen presentation to CD4+ and CD8+ T lymphocytes. The largest genome-wide association studies to date by the International Multiple Sclerosis Genetics Consortium (IMSGC) discovered 110 common genetic variants associated with MS, most of which were enriched in inflammatory pathways[17]. In line with these genetic findings, histopathological studies have found the infiltration of mononuclear phagocytes, T lymphocytes, B lymphocytes, plasma cells, and dendritic cells in acute white matter lesions, which are important diagnostic MRI markers for MS[18]. In the last 20 years, the use of immunotherapies that target these immune system components have been demonstrated to be particularly effective in the early stage of MS for reducing relapse rates and disability progression. However, in the progressive phase of the disease, immunotherapeutics have had little or no clinical effects[19], suggesting that as the disease activity progresses, the inflammatory process decreases, and other mechanisms drive the underlying disease processes.

1.4.1 The innate immune response

There are two key aspects to the immune system, the innate and adaptive immune responses, both of which are involved in the development of MS[20]. The innate, or

nonspecific, immune response is an evolutionarily older defense mechanism that acts nonspecifically against pathogens. Although many cells are involved in the innate immune responses in the CNS, the predominant cells involved in the pathological process are macrophages and microglia[20]. Macrophages, which are generated through a series of changes after the circulating blood monocyte enters the tissue, can not only cause tissue damage through direct neuroinflammatory processes but are also involved in the remyelination process that repairs the tissue[21]. Work in the EAE model suggests that the deleterious macrophages are primarily differentiated from inflammatory monocytes, whereas the remyelinating macrophages are derived from patrolling monocytes[22]. A similar dichotomy is observed for microglia, which not only play an important role during CNS inflammation but also seem to act in CNS surveillance to maintain local environmental stability[23,24].

1.4.2 The adaptive immune response

Unlike the innate immune system, the adaptive immune system is highly specific to particular epitopes, with antigen-specific receptors on T-cells and B-cells[20]. T-cells mature in the thymus and consist of two major subsets—CD4+ helper T-cells and CD8+ cytotoxic T-cells[25]. In the development process, antigen-binding domains are reshuffled into forms that can bind virtually any structural epitope, including those that comprise host cell structures. The potential negative impacts of this type of recognition are addressed in the process of thymic negative T-cell selection, resulting in the induction of T-cell apoptosis after the binding of the host self-epitopes and therefore preventing the subsequent possible appearance of an autoimmune response. An autologous selection process occurs in the bone marrow during B-cell development to eliminate the creation of self-reactive autoimmune B-cells. However,

because the selection process is sometimes incomplete, some autoreactive T- and B-cells can persist in the peripheral immune system. These autoreactive cells may never be activated. However, they maintain the potential to be activated by encountering a self-antigen that binds to their T- or B-cell receptor or a sufficiently similar pathogenic epitope via the mechanisms of molecular mimicry, or they can even be activated via bystander activation, potentially resulting in autoimmune disease[26].

CD4+ T-cells can be differentiated into four major subsets: Th1, Th2, regulatory T-cells (Treg), and Th17. Th1 cells secrete pro-inflammatory cytokines, including TNF-α, interleukin-2 (IL-2), and the prototypical Th1 cytokine IFN-γ, all of which induce a cell-mediated immune response against intracellular pathogens that does not involve antibodies[27]. Th2 cells produce IL-4, IL-5, IL-13, and IL-10, among others, which stimulate a humoral immune response against extracellular pathogens that involve antibodies [27]. Treg cells produce regulatory cytokines, particularly TGF-β, which downregulates the immune response, maintaining homeostasis between immune activation and immune suppression[28]. It has been shown that TGF-β prevents Th1 and Th2 differentiation and initiates Th17 differentiation in combination with IL-6. The Th17 cells secrete lineage-specific factor RAR related orphan receptor C (RORC) and have been identified as a distinct linage of CD4+ T-cells[29]; they also produce pro-inflammatory cytokines including IL-17, IL-21, IL-22, and IL-26[30,31], typically during more long-term immune responses.

In the CNS, antigens presented by the cells (microglia and B-cells, etc.) expressing HLA class II molecules can be recognized by CD4+ T-cells, whereas antigens presented by HLA class I molecules can be recognized by CD8+ T-cells. In MS, the clonal expansion of both CD4+ T-cells and CD8+ T-cells can be observed in brain

lesions, suggesting that antigen-specific T-cell responses contribute to the development of the pathology[32,33].

1.4.3 The role of the adaptive immune system in the induction of MS

Despite the emerging consensus regarding the involvement of the adaptive immune system in the development of MS, its role in the induction of the early inflammatory phase and the progressive phase of MS remains controversial; specifically, it is not known which process happens first. One hypothesis suggests that the first event—the activation of a CNS antigen-specific immune response—may take place in the periphery, followed by the invasion of autoreactive immune cells into the CNS. This hypothesis is supported by most findings from EAE models with similar processes. In MS, as a result of molecular mimicry, bystander activation, or viral persistence with or without epitope spreading[26], autoreactive T-cells and B-cells are activated in the peripheral tissues, which then migrate to the lymph nodes and ultimately invade the CNS. The cytokines and antibodies released by the cells can target the myelin sheath and glial cells, resulting in the demyelination process that underlies the earliest phases of the disease. During the chronic inflammation phase, the unbalanced demyelination/remyelination process can lead to the formation of lesions[18]. The inflammation process continues to drive the progression of the disease. In the lesions of patients with progressive disease, the immune cell infiltration correlates with axonal injury, and the aggregation of B-cells in the meninges is associated with the increasing severity and clinical manifestations of the disease[34].

However, this hypothesis has been subject to criticism because the loss of oligodendrocytes and the activation of microglia can be observed in the normal-appearing white matter without the infiltration or presence of lymphocytes[18].

Therefore, another hypothesis has been proposed—that the initial pathological event for MS is the neurodegenerative process, which is followed by a secondary activation of innate and adaptive immune responses. Supporting this hypothesis, the currently used immunomodulatory interventions have limited success in changing the disease course for progressive MS[35-37]. Despite the apparent conflict, it should be noted that both processes are not mutually exclusive; both intra- and extra-CNS immune activation may act synergistically to contribute to the development of MS.

1.5 Environmental factors in the development of MS

The uneven distribution of MS by geography and the change in risk as people migrate suggest that environmental factors acting at the personal and population levels have a role in the development of MS[38,39]. Although the environmental factors involved have not all been identified, extensive studies support three environmental risk factors in MS etiology: infection with human herpesviruses, particularly Epstein-Barr virus (EBV) and human herpesvirus 6 (HHV-6); vitamin D deficiency; and cigarette smoking[38-40]. These factors can exert effects long before MS becomes clinically evident.

1.5.1 Infection with Epstein-Barr virus (EBV)

EBV, also called human herpesvirus 4 (HHV-4), is a ubiquitous double-stranded DNA virus with a diameter ranging from 122 to 180 nm that is covered by nucleocapsid, viral tegument, and envelope[41]. After the acute infection of B-cells and epithelial cells, EBV undergoes lytic replication and produces infectious virions[42]. It then goes into a latency phase—Latency I, Latency II, or Latency III, depending on the cell type—in which only a portion of the EBV genes are expressed. In this stage,

the EBV genome exists as an episome in the host cell's nucleus and is copied by cellular DNA polymerase[42]. A summary of the key viral antigens, which are expressed during the lytic and latent stages and can be used for serological tests for EBV infection, is shown in Table 1.1.

Table 1.1: Summary of the key EBV viral antigens.

Protein/antigen	Stage	Description
EBNA-1	latent+lytic	The first EBV protein detected and the only viral protein expressed in all forms of latency
EBNA-2	latent+lytic	EBNA-2 is the main viral transactivator
EBV-EA	lytic	early antigen
EBV-VCA	lytic	viral capsid antigen

EBV infection is common in the general population (~90%), but its frequency is even higher among persons with MS, typically approaching 100%[43]. The first evidence that EBV might play a role in the pathogenesis of MS was reported in 1979 when Fraser and colleagues found that peripheral blood lymphocytes from MS patients had an increased tendency to transform to EBV-induced B lymphocytes *in vitro*[44]. In a nested case-control study involving over 3 million US military personnel, Levin and colleagues found that, before the age of 20, the mean anti-EBNA IgG antibody titers were similar for controls and individuals who in the future would develop MS. However, a remarkable increase in anti-EBNA antibody titers was observed for individuals who later developed MS, while the level for controls remained constant[45]. Levin and colleagues also found that the risk for developing MS was extremely low for individuals with no EBV infection, but the risk increased sharply for those individuals after EBV infection[46]. The consistency of the association of anti-EBNA IgG seropositivity in developing MS was validated in a recent meta-analysis[47].

To explain the mechanisms by which EBV might contribute to the development of MS, four hypotheses have been proposed. In the molecular mimicry hypothesis, activated T-cells after EBV infection could cross-react and attack EBV antigens and CNS antigens because both antigens share similar epitopes[48,49]. In the bystander damage hypothesis, the direct and specific immune response against EBV infection and/or reactivation results in bystander damage to the CNS[50]. The αB crystalline hypothesis posits that the small heat shock protein αB crystalllin, which is derived from oligodendrocytes, is expressed after EBV infection. Then, the immune system mistakenly targets the αB crystalllin by generating a CD4+ T-cell response, resulting in demyelination[51]. In the fourth hypothesis, the EBV-infected autoreactive B-cell hypothesis, EBV-infected autoreactive B-cells infiltrate the CNS, clonally expand in the target site, and activate autoreactive T-cells, initiating an autoimmune attack on the CNS[52].

1.5.2 Infection with HHV-6

HHV-6, which has two variants (HHV-6A and HHV-6B) and shares some genetic attributes with EBV, was first discovered in 1986[53]. Both variants are double-stranded DNA viruses; the HHV-6A viral genome is 159 kb long and contains 119 unique open reading frames (ORFs), whereas the HHV-6B viral genome is 162 kb with 119 unique ORFs[54,55]. Their genomic architectures can be categorized into two major regions: the unique long region (~143 kb) consists of the genes involved in replication, cleavage, and viral genome packaging, and the short region (~8 kb) contains the genes responsible for viral integration into host chromosomes[55]. The genome is packaged within a capsid surrounded by tegument and then encased in a lipid envelope derived from the host cell membrane[56].

HHV-6 infection is highly prevalent (>90%) in the general population; the primary infection occurs in early childhood, which then establishes a latent infection within peripheral blood mononuclear cells (PBMCs)[57] and saliva[58]. The ORF U94 transcript is the only molecular marker that can be detected during latency and may control the establishment and maintenance of these processes[56].

The first study that provided evidence supporting the potential role of HHV-6 in MS was performed by Challoner and colleagues[59]. The authors found the presence of anti-HHV-6 monoclonal antibodies specific to virion proteins in MS-derived oligodendrocytes, but these antibodies were absent in control individuals. In addition, in some cases, the presence of these antibodies was located around MS plaques and not within the surrounding CNS white matter. However, the role of HHV-6 in MS remained controversial, and subsequent studies reported contrasting results. One study showed significantly higher anti-HHV-6 IgM in MS sera compared to controls,[60] whereas others reported similar titers of anti-HHV6-IgM and IgG between MS and controls[61]. In PBMCs from MS patients, Chapenko and colleagues[62] observed active HHV-6 mRNA transcription. However, Rotala and colleagues[63] did not observe HHV-6 mRNA expression; one explanation may be that all of the patients in their study were at advanced stages of the disease and received anti-inflammatory medication. Interestingly, the authors performed a second study on patients[64] with early stages of MS with no medication and found that the HHV-6 virus was replicating in various parts of the CNS closely associated with active demyelinating regions.

In longitudinal studies, consistent results show that increased anti-HHV-6 IgG titers predict a worse clinical outcome (relapses and disability) in MS patients[65-67]. A

potential explanation the involvement of HHV6 in MS is molecular mimicry because HHV-6 shares significant sequence homology (HHV-6 U24, residues 4-10) with myelin basic protein (MBP, residues 96-102). This homology is relevant, given the key role of MBP in maintaining the integrity of the myelin sheath; MBP dysfunction can result in the development of demyelinating diseases such as MS[68]. Research has also found that the level of MBP-reactive T-cells is elevated in MS, potentially allowing these cells to recognize and initiate an immune response towards both MBP and HHV-6[68].

1.5.3 Vitamin D deficiency

The geoepidemiology of MS correlates not only with age at EBV infection but also with winter ultraviolet radiation (UVR) exposure and vitamin D deficiency[40,69]. In individuals of European descent, vitamin D is predominantly produced upon the photolysis of precursor sterols in the skin following exposure to UV, whereas dietary vitamin D intake is a relatively minor contributor[70,71]. The biologically active form of vitamin D is 1,25-dihydroxyvitamin D ($1,25(OH)_2D$), and the major circulating and diagnostic form is 25-hydroxyvitamin D (25(OH)D), the optimal levels of which are disputed but thought to be at least 75 nmol/L for immunomodulatory functions[71]. In a nested case-control study, non-Hispanic white individuals in the highest quintile of 25(OH)D levels (>99 nmol/L) showed a 62% lower risk for developing MS compared to those in the lowest quintile (<66 nmol/L)[72]. The mechanism by which vitamin D exerts a protective effect on MS is unclear, but its immunomodulatory effects and/or the actions of the ubiquitous vitamin D receptor, which functions in the regulation of many genes, including those found in nervous and immune system cells, are thought to be involved.

1.5.4 Cigarette smoking

In addition to the diverse negative impacts on health of tobacco use and particularly smoking, smoking is suggested to have a role in MS onset and progression. Unlike EBV infection and vitamin D/UV deficiency, cigarette smoking cannot explain the uneven distribution of MS by geography and the altered risk upon migration. However, the evidence supports smoking as an important risk factor in MS[38].

MS risk in ever-smokers is approximately 50% higher than never-smokers, and studies have shown a direct association with the duration and intensity of smoking[73,74]. Possible mechanisms that may explain the effects of smoking on MS risk include the disruption of the blood-brain barrier by nicotine[75], the alteration of normal immune function, and inflammation induced by smoking[76]. Smoking has been recognized to elevate important markers of inflammation and autoimmunity such as C-reactive protein, intercellular adhesion molecule-1, and E-selectin[76]. Furthermore, cigarette smoke contains an array of potent oncogenic compounds, and thus an impact on DNA and RNA is highly probable.

1.6 Genetic factors associated with MS

The earliest evidence that genetic factors might be involved in the development of MS came from the findings of familial aggregation and MS risk variation among ethnic groups living in the same geographical regions. Compared with the age-adjusted risk in the general population, the recurrence risk increases to 1% for second degree and third degree relatives and further increases to 3% in first degree relatives[77-79]. Twin studies have consistently shown that the concordance rate in monozygotic twins (25%) is five times higher than for dizygotic pairs[80,81]. Among first degree but non-

biological relatives (adopted) living with the index case, the risk for MS was significantly lower than for biological relatives and not clearly different from the overall population risk[82]. The same is true for offspring recurrence rates, in which children of one MS affected parent had lower risk than when both parents had MS[83]. Taken together, this evidence suggests that in addition to environmental exposures, genetic factors also influence an individual's susceptibility to MS. The advances in identifying the genetic factors derive primarily from linkage analysis and genome-wide association analysis, which we will explore below.

1.6.1 Linkage analysis

The concept of linkage analysis is based on the observation that genes located physically close to each other on a chromosome remain linked during meiosis[84]. In the latter half of the twentieth century, it was the primary mode for statistical genetic mapping of Mendelian and complex traits with familial aggregation[85]. A logarithm of odds (LOD) score ≥3, which is computed as the logarithm of the odds that the loci are linked divided by the odds that the loci are unlinked, is generally considered evidence of linkage.

Despite the success of linkage analysis in other neurodegenerative diseases such as Parkinson's disease[84], it has been less successful in MS. Unlike other complex diseases, it is rare to find more than three or four affected individuals within families; it is also thought that the existence of rare, highly penetrant alleles in MS is uncommon if it occurs at all[86]. In late 2005, the IMSGC performed the largest non-parametric genome-wide linkage screen in 730 multiple MS families. They found significant linkage within the *HLA* locus with a peak LOD score of 11.7, whereas none of the loci outside the *HLA* locus reached statistical significance for linkage[87].

1.6.2 Genome-wide association studies (GWAS)

Given that common variants have relatively modest effects on the development of complex diseases such as MS (OR ranging from ~1.05 to 1.30 for non-HLA loci[17]), studies have shifted from linkage analysis to association analysis. Accordingly, GWAS utilizing the genomic variation marker, the single nucleotide polymorphism (SNP), have become the preferred mapping tool[88]. To increase efficiency and cost, GWAS typically use a multistage design. In stage one, a full set of SNPs is genotyped in a relatively small number of samples, and a subset of SNPs with a liberal p-value threshold are then selected for follow-up. In stage two and beyond, as applicable, the SNPs selected from stage one are genotyped in a larger sample size. Finally, p-values from the different stages are combined, and a p-value of 5.0×10^{-8} (Bonferroni correction of p-value 0.05 for 1 million independent tests) is widely used as the conservative genome-wide significance threshold[89].

Since the first GWAS publication in 2005 and continuing until 2016, GWAS have successfully identified over 1,800 loci that confer susceptibility to more than 600 complex diseases[90]. However, there are some major challenges for GWAS. The first is population stratification, in which systematic differences of allele frequencies between cases and controls are due to different ancestry[91]. If not properly corrected, this can lead to an increasing number of false positive results. A common method for measuring population stratification is the calculation of the genomic control (λ_{GC}) [92-94], which is defined as the median χ^2 (1 degree of freedom) association statistic across SNPs divided by its theoretical median under the null distribution. If $\lambda_{GC} = 1$, this suggests no population stratification is present, whereas $\lambda_{GC} > 1$ suggests possible population stratification is present or may indicate the presence of other confounders

such as cryptic relatedness[95]. In GWAS, a $\lambda_{GC}<1.05$ is generally considered to be acceptable for presenting association results[91]. To adjust for population stratification, methods such as principal components analysis and multidimensional scaling analysis have been used. However, these methods cannot account for other confounders such as cryptic relatedness.

Another major challenge is deriving the multiple testing thresholds, because thousands to millions of tests are conducted. Currently, the simplest and most widely used method is Bonferroni correction. However, this method can sometimes be excessively conservative because some markers are in linkage disequilibrium (LD), leaving the effective number of independent tests (N_{eff}) much lower than the actual number of tests conducted (N). Many researchers[96] have proposed algorithms to establish the correlation among tests and calculate the N_{eff}, upon which N is replaced by N_{eff} in the corresponding formula of Bonferroni correction. Another widely used correction method is to determine the false discovery rate (FDR)[97], which estimates the proportion of false positives from the significant results ($\alpha=0.05$). Under the null hypothesis that there are no true associations for the GWAS, the p-values of the associations would follow a uniform distribution from 0 to 1. The gold standard correction method used in GWAS is permutation testing[98], although it can be computationally intensive. Permutation testing generates the empirical distribution of test statistics for a specified dataset assuming the null hypothesis is true. It then randomly reassigns the dataset samples' phenotypes, while the genotypes are kept the same. A predefined number of replications are repeated to generate an empirical distribution with resolution equal to the number of replication cycles. These methods have already been incorporated into genetics analysis tools such as Plink[99].

1.6.3 GWAS in MS

The first MS GWAS was published in 2007 by the IMSGC, and it identified associations with variants in both *IL7R* and *IL2RA3*[100]. Since this first MS GWAS publication, only 15 genome-wide association studies have been published[86,101]. The susceptibility role of the *HLA* region has been replicated in all 15 GWAS analysis, but the successful identification of new risk loci has required a large number of subjects. The first GWAS (sample size <5,000) only identified two non-HLA SNPs, whereas the latest GWAS published by the IMSGC in 2013 included 29,300 cases and 104,185 controls and identified 97 statistically independent non-HLA SNPs[17].

1.6.3.1 The HLA region in MS

The extended HLA gene complex, which primarily includes HLA class I, class II, and class III genes, and the extended region, is located on the short arm of chromosome 6 and spans a region of 7.6 Mb[102]. The HLA class I region includes classical class I genes (HLA-A, -B, and -C), genes encoding the MHC I proteins that present endogenous antigens to CD8+ T-cells. The HLA class II region includes classical class II genes (HLA-DP, -DQ, and -DR), genes encoding the MHC II proteins that present exogenous antigens to CD4+ T-cells. The HLA class III region includes the genes encoding components of the complement system[102]. The gene map of the extended HLA region is shown in Figure 1.5.

Figure 1.5: Gene map of the extended HLA region (reproduced from Horton and colleagues[102])

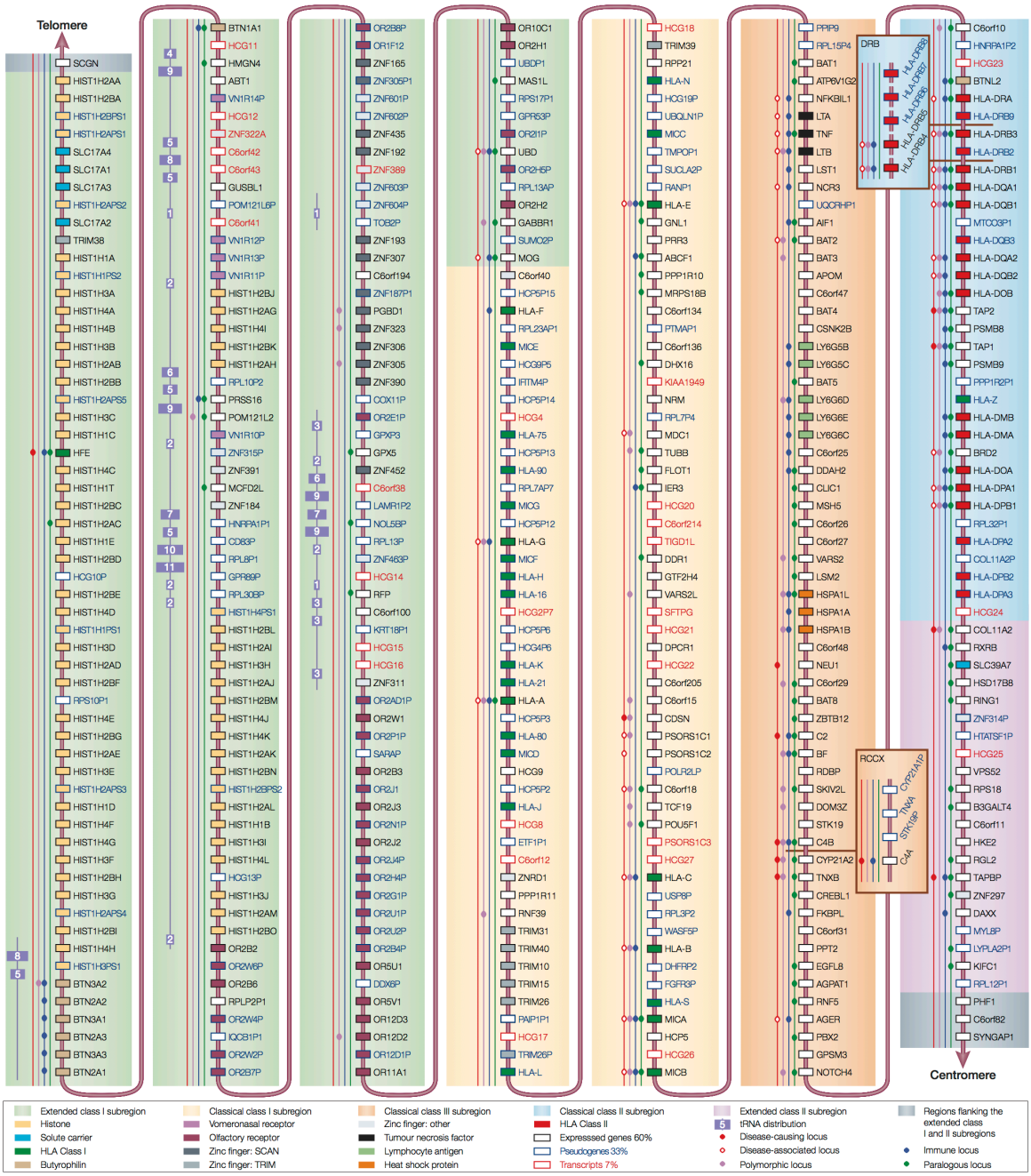


Figure | **Gene map of the extended major histocompatibility complex (xMHC).** The gene map is shown from telomere (left) to centromere (right) on the short arm of chromosome 6. The five colour-coded subregions making up the xMHC span about 7.6 Mb and are defined as: the extended class I subregion (green block; *HIST1H2AA* to *MOG*; 3.9 Mb), the classical class I subregion (yellow block; *C6orf40* to *MICB*; 1.9 Mb), the classical class III subregion (orange block; *PPIP9* to *NOTCH4*; 0.7 Mb), the classical class II subregion (blue block; *C6orf10* to *HCG24*; 0.9 Mb) and the extended class II subregion (pink block; *COL11A2* to *RPL12P1*; 0.2 Mb). Regions that flank the extended class I and II subregions are shown as grey blocks. Insets denote the hypervariable RCCX and DRB regions. Numbers and positions of tRNA genes are represented by indigo bars, the length of which is proportional to the gene number between other loci. Vertical lines connect the two main groupings of tRNA genes of 1.6 Mb and 0.5 Mb of the sequence (separated by 0.6 Mb). Circles to the left of each locus indicate disease status, polymorphism, immune status and paralogy as described in the text. The gene map of the xMHC is also available as a poster, which accompanies this issue. The poster is available online (http://www.nature.com/nrg/journal/v5/n12/poster/MHCmap).

The association of the HLA locus as a major susceptibility region for MS has been validated in all populations and in different subtypes of MS[86]. However, due to the strong linkage disequilibrium within the region[103], it is hard to identify true independent risk loci. The thorough work by Patsoupolous and colleagues utilized imputation methods to investigate the independent effects in the HLA region using 5,091 MS and 9,595 healthy controls[104]. They found that the most significant variant was *HLA-DRB1*15:01* (OR=2.92), which is in agreement with other published results[16,87]. They also performed a conditional analysis by adding the top significant variants as model covariates and identified another nine statistically independent effects within the extended HLA region, including the following: five *HLA-DRB1* (*HLA-DRB1*03:01*, *HLA-DRB1*13:03*, *HLA-DRB1*04:04*, *HLA-DRB1*04:01*, and *HLA-DRB1*14:01*) and one *HLA-DPβ1* (*rs9277489/Leu65*) in class II and one *HLA-A* (*HLA-A*02:01*) and two *HLA-B* (*HLA-B*37:01* and *HLA-B*38:01*) in class I. Cumulatively, all these independent effects accounted for 14.2% of the total variance in MS susceptibility.

Functional studies have suggested a possible functional role for *HLA-DRB1* in susceptibility to MS. The shape and charge of the P4 pocket in the HLA molecule containing *HLA-DRB1*15:01* is distinct from other *HLA-DRB1* molecules (Figure 1.6). Such a difference would lead to high affinity binding to the MS autoantigen MBP and potentially result in a more severe inflammatory response[105,106].

Figure 1.6: Molecular modeling of HLA-DRB1 susceptibility alleles and resistance alleles in MS, reproduced from Oksenberg and colleagues[106]

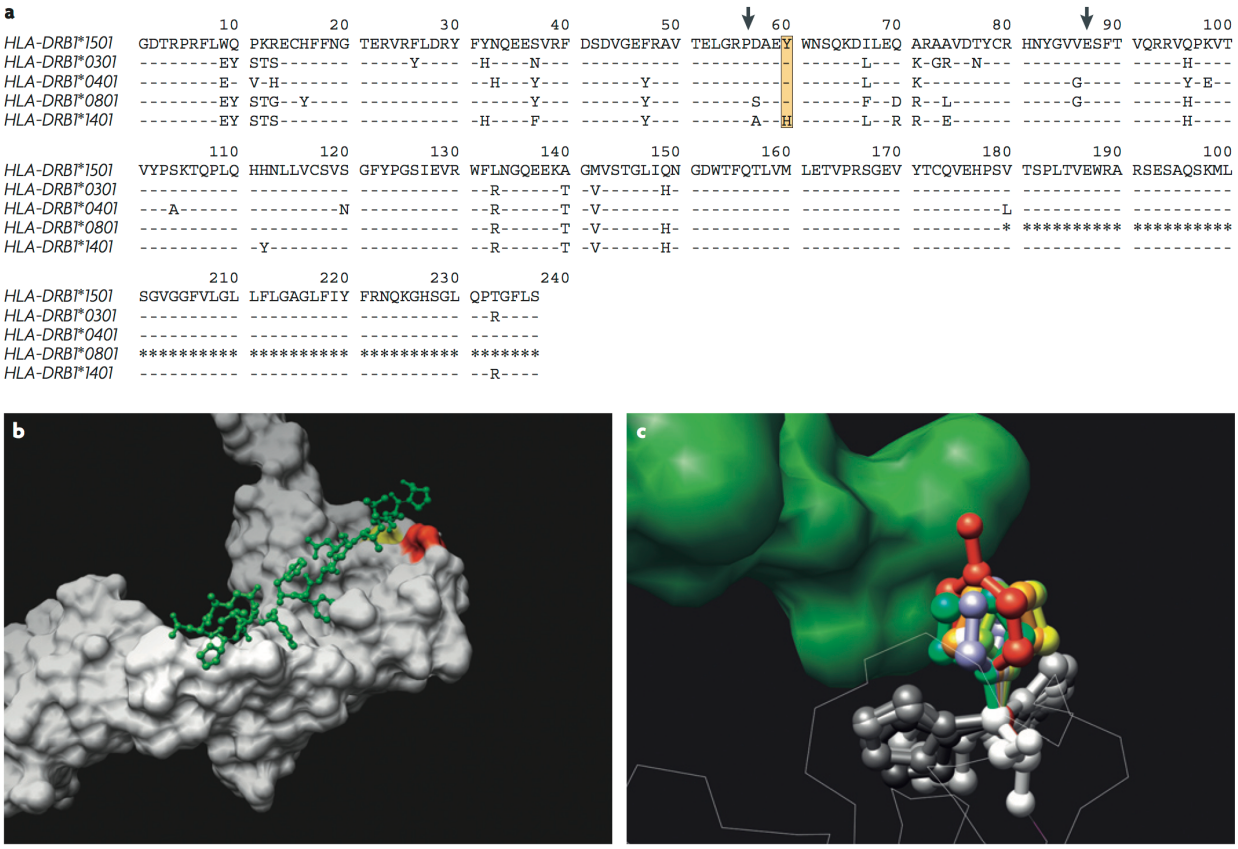


Figure: **Molecular modelling of *HLA-DRB1* susceptibility alleles and resistance alleles in multiple sclerosis.** **a** | Sequence alignment of various HLA-DRB1 alleles compared with HLA-DRB1*1501. Arrowheads indicate the sequence of the floor of the groove (peptides 56–87) that mostly interacts with myelin basic protein (MBP) peptides 85–98. Whereas all permissive alleles exhibit a tyrosine (Y) in position 60 (boxed), the resistance allele DRB1*1401 has a histidine (H). The alignment was performed using the Immunogenetics (IMGT) database. **b, c** | Crystallographic structure of HLA-DRB1*1501 bound to MBP peptides 85–98. **b** | Side view of the complex with the HLA molecule (white) in surface view and MBP peptides 85–98 (green) in ball and stick. Allele discriminatory position β60 (Y or H) is coloured in red to highlight its proximity to the peptide, potentially altering its binding properties. Residue βA71, which is crucial for peptide binding, is shown in yellow. The small and uncharged βA71 creates the necessary room for the binding of the large hydrophobic side antigen chain in the P4 pocket. **c** | Close-up view of HLA-DRB1 position β60 showing the Y residue (red) of allele HLA-DRB1*1501, and all possible rotamers of a modelled Y to H substitution. The OH^- terminus of the MBP peptide (threonine 97 and proline 98) is shown in green in the ball and stick structure. In the absence of a crystal structure for HLA-DRB*1401, all possible rotamers of a Y–H substitution were modelled using Swiss-PDB viewer. The most thermodynamically favourable rotamers are displayed in colours, whereas the least favourable ones are show in grey. Visualizations were performed with the software Chimera.

1.6.3.2 The 110 non-HLA established variants associated with MS

To date, GWAS studies have identified 110 non-MHC susceptibility variants, showing moderate effects with ORs ranging from ~1.05 to 1.30. Of the 110 SNPs, 15 are coding variants, whereas 35 are in high LD with coding variants according to the

prediction tool Variant Effect Predictor[107]. However, upon further prediction of the damaging effects of these coding variants using PolyPhen[108] and SIFT[109], only 14 are missense, and 7 are possibly damaging to the transcriptional product. By checking whether the SNPs are quantitative trait loci for expression (eQTL) using an eQTL database (University of Chicago), researchers found that 15 SNPs are eQTLs and that 16 SNPs are in LD with SNPs that are eQTLs[86]. The IMSGC has undertaken a further GWAS combining all genotyping MS experiments to date from approximately 40,000 cases and 40,000 controls and has increased the number of associated SNPs to 193 (manuscript under preparation).

The primary role of these established MS risk loci has been in the T-cell-mediated inflammation process[16]. Although inflammation itself has been causally linked to neurodegeneration, risk genes that are immune independent and primarily act in a neurodegenerative role remain largely undiscovered. It is likely that these immune-independent risk genes have not been discovered because almost all MS GWAS have used a case-control study design, thereby necessarily focusing on initiating events (the first hallmark of MS, the inflammation process in greater than 85% of cases) rather than severity and the neurodegenerative pathways involved in MS progression[110]. Of these studies, only three GWAS[101,111,112] have focused on MS severity. The results were not consistent, although two of the studies[111,112] demonstrated a significant association of *GRIN2A,* which encodes a member of the glutamate-gated ion channel protein family, with the protein products that affect the efficiency of synaptic transmission and are possibly involved in memory and learning. Larger studies that focus on MS severity must be performed to confirm these findings and potentially identify other novel genes.

For MS, efforts to identify susceptibility loci that predict the MS clinical course (relapse, disability progression) are still in their comparative infancy relative to onset-associated genetic studies. Recently, we investigated the association between 52 non-MHC MS risk alleles[16] and disease severity in a large meta-analysis that included 7,125 MS cases. A weighted genetic risk score (wGRS) was calculated for each MS patient using the log value of the discovery odds ratios as the weight for each of the 52 non-MHC risk alleles. The number of risk alleles for each SNP was multiplied by the weight for that variant, and then the sum across all 52 variants was calculated. This wGRS was not strongly associated with disease severity after adjusting for cohort, gender, age of onset, and *HLA-DRB1*15:01* genotype. After restricting the analysis to cases with disease duration ≥ 10 years, all associations were null (p-value all ≥ 0.50). No single SNP was associated with disease severity after adjusting for multiple testing[113].

1.6.4 The missing heritability in MS

Cumulatively, however, these loci (110 non-HLA + HLA) can only explain approximately 28% of the perceived heritability of MS[17]. The exact component of MS risk that is heritable is difficult to determine due to the impact of environmental factors, but estimates based on twin studies and MS families suggest that the risk may be between 15 and 40%[2]. Therefore, our current level of genetic knowledge can explain in the region of 5 – 8% of the risk of MS, depending on ethnicity and environment.

This leads to the following question: where is the missing heritability? The strategy we discussed previously, which is to focus on SNPs that predict clinical course using a longitudinal study design, may be able to explain part of the missing heritability

problem in MS. Other hypotheses include lower frequency variants and rare variants, gene-gene interaction, gene-environmental interaction, and epigenetics.

1.6.4.1 Lower frequency variants and rare variants

A current popular hypothesis suggests the presence of lower frequency variants (0.5% < MAF (minor allele frequency) < 5%) and rare variants (MAF < 0.5%) that exert a much larger effect (OR > 2) may account for a large fraction of this missing heritability[114]. A GWAS based on the "common disease-common variants hypothesis", seeking to examine SNPs with MAF > 5% that exert small to modest effects (OR<1.3), could not accurately capture the lower frequency/rare variants. Thus, researchers now favor a shift toward the use of next generation sequencing (NGS) technologies to identify these rarer variants, either by sequencing the extended genomic regions strongly indicated in GWAS or simply sequencing the whole genome[115]. However, attempts to discover relevant rare variants have not been as fruitful as expected. The first whole-exome sequencing in MS was published in 2011 and included the sequencing of forty-three individuals with MS (1 from each family). A rare variant in the *CYP27B1* gene that results in the complete loss of gene function was identified[116]. However, in subsequent independent replication studies, this result could not be confirmed[117-119]. Recently, Hunt and colleagues performed sequencing for the coding exons of 25 GWAS risk genes in 24,892 subjects with six autoimmune disease phenotypes (MS included) and 17,019 controls. Their results did not support the hypothesis that unobserved rare causal variants resulted in significant associations at common tag variants. Instead, they concluded that rare coding region variants at known loci have a negligible role in common autoimmune disease susceptibility[120]. The 1000 Genomes Project has revealed that on average, each person carries some

proportion of rare variants, but their role in the development of disease could not be presumed if no other functional studies were performed to validate them[121]. Therefore, other researchers have argued that rare variants may be unlikely to contribute significantly to the heritability in MS. Although controversial, this will remain a hot area of study in the near future. Interestingly, the IMSGC has presented in poster form the preliminary results of its large exome array study in MS. This experiment has been extremely difficult and has thus far yielded only a small number of potential rare variant targets that do not significantly add to the heritability of MS, suggesting that this hypothesis is unlikely to explain the bulk of missing heritability.

1.6.4.2 Gene-gene interaction

Thus far, GWAS have been successful in identifying susceptibility loci that exhibit additive and independent effects on traits using a single locus analysis strategy. However, for complex diseases such as MS, where genes may not act in isolation but interact with each other through complex mechanisms, suggests that analyses that take into account gene-gene interactions may yield greater effects on MS risk and clinical course. Therefore, analyzing the statistical gene-gene interactions, which are defined as the interaction variance explained by the combinations of causal variants that were not due to their independent effects, may explain a portion of the missing heritability of MS[122,123].

Regression models were widely used to test for the interaction of two or more genetic loci that contribute to disease risk[123]. For binary outcomes (cases and controls), the most straightforward approach to analyze the two loci interactions was to define two logistic regression models, one that takes into account only marginal effects of both SNPs and another with a full logistic regression that includes pairwise interaction

between the two SNPs. A likelihood ratio test with four degrees of freedom can then be used to test whether there is a difference between the two models[124]. However, these methods have some challenges to exhaustively search for interactions using GWAS data, because the number of tests would exponentially increase with the number of SNPs included, which is not only time consuming but also raises significant multiple testing issues[123].

Therefore, in MS, the reported gene-gene interaction results were largely based on the plausible candidate gene /pathway approach or by selecting a liberal p-value (for example, $p < 1.0x10^{-4}$) for the SNPs in a GWAS discovery stage and then testing the interactions between them. The most commonly identified gene-gene interactions were genes within the HLA region[125-129] or non-HLA genes that interact with HLA genes[130,131]. For HLA gene interactions, Lincoln and colleagues found interactions between *HLA-DQA1*0102* and *HLA-DRB1*15* within the HLA class II region, in which HLA-*DQA1*0102* strongly increased MS risk in the presence of *HLA-DRB1*15* while showing a protective role in its absence[125]. The interaction of *HLA-DRB1*15* with *HLA-DRB1*01* not only predicts MS risk[126,127] but also influences the clinical course of MS[128]. The latest work by the IMSGC, which used 17,465 and 30,385 controls to further substantiate the class II HLA interactions, found the following two new interactions: *HLA-DQA1*01:01-HLA-DRB1*15:01* and *HLA-DQB1*03:01-HLA-DQB1*03:02*[129]. For interactions between HLA and non-HLA genes, the ANZgene consortium[130] found a statistical interaction for SNPs within the non-HLA gene *EVI5* and *HLA-DR15*; these results were further validated by Johnson and colleagues[131]. For non-HLA gene interactions, using a knowledge-driven interaction analysis, Bush and colleagues[132] found significant interactions for genes

enriched in calcium signaled cytoskeletal regulation, in which *CHRM3* interacts with *MYLK*, *PLCβ1* interacts with *PLCβ4*, and *ACTN1* interacts with *MYH9*.

1.6.4.3 Gene-environmental interaction

Recognition and emphasis is growing that gene-environmental interactions, which are defined as the joint effects of genes and environmental factors that cannot be explained based on their independent marginal effects, also accounted for a part of the missing heritability of MS[133,134]. Two primary models are used to investigate this form of interaction. One interaction model is the multiplicative model, in which the joint effects of genetic and environmental factors on the relative risk of disease differs with the product of the relative risks of each factor separately. Another interaction model is the additive model, in which the joint effect differs with the sum of the relative risks of each factor separately[133].

To test and identify gene-environmental interactions, the widely used study design is the case-control study, in which genetic factors and environmental exposures are measured with the recruitment of disease cases and samples of disease-free controls[135]. In a population-based case-control study, van der Mei and colleagues found a significant interaction between HLA-DR15 and infant sibling exposure in MS; the combined effect of HLA-DR15 positivity and low infant sibling exposure (cumulative exposure to younger infant siblings (< 1 year) before the subjects were aged 6 years) 3.9 fold greater than expected[136]. To further explicitly evaluate the relationship of the combined effects of having known risk factors (*HLA-DR15* genotype, the prior acquisition of infectious mononucleosis, high anti-EBV nuclear antigen IgG status, low lifetime sun exposure, low 25(OH)D level, and having ever smoked), they performed an incident case-control study using cases with a first

clinical diagnosis of central nervous system demyelination (FCD) and population-based controls. They discovered a significant interaction on the additive scale between *HLA-DR15* and a history of infectious mononucleosis in predicting FCD onset. The five known risk factors accounted for 63.8% of the FCD onset, mostly driven by *HLA-DR15* genotype, smoking, and lifetime sun exposure[137].

However, results produced using the above study design can only suggest association but not causality, because this design typically retrospectively investigates for evidence of interactions that existed before disease onset. The important advantage of the case-control design is in the study of rare diseases, because it is relatively easier to recruit these samples through specialized centers that identify and diagnose the disease. However, case-control studies are subject to significant sources of bias, for example, survival bias, diagnosis bias, surveillance bias, etc.[135,138] By contrast, the prospective cohort study design, which investigates representative samples of the population before disease onset and follows until the specified endpoints occur, significantly reduces the weaknesses and biases of a case-control design. The results derived from this design can provide evidence for causality because it identifies the factors that predispose an individual to the development of a disease in the latter follow-up period. The prospective study standardizes and details the collection of exposure information and follows all participants in a systematic way. Therefore, the environmental factors, especially with respect to gene-environmental interactions, can be more precisely assessed, and this study design is less prone to bias compared with case-control designs[135].

Compared to the more commonly published gene-environmental interactions in predicting MS using a case-control design (detail reviewed in[134]), the use of

prospective studies assessing the gene-environmental factors in predicting the MS clinical course is in its infancy because these studies typically require many years of follow up. In work by our group using a prospective cohort of 141 participants with RRMS and followed between 2002-2005, Rui and colleagues found a significant interaction between rs2248359 in *CYP24A1* and the 25(OH)D level in predicting relapse; carriers of the CC genotype showed protection for relapse with increased levels of 25(OH)D[139]. In another study using the same cohort, two novel genes (CYP2RA and PRKCB) were identified that interact with 25(OH)D levels in predicting relapse[140]. In chapters 3, 4, and 5, I will discuss in detail our novel findings using such a strategy to investigate the gene-environment interactions in predicting the MS clinical course.

1.6.5 Epigenetic modifications

Epigenetics refers to a functionally relevant changes of gene expression without permanent modification of the DNA sequence[141]. In MS, epigenetic changes can mediate the response to environmental signals (e.g., EBV infection, vitamin D deficiency, and smoking)[38,39], which can then also influence gene expression. Epigenetic mechanisms include DNA methylation, the modification of nucleosomal organization, and miRNA-associated post-transcriptional gene silencing [142].

DNA methylation is the most extensively studied epigenetic mechanism. Methylation predominantly occurs by adding methyl groups ($-CH_3$) at carbon-5 of cytosine amino acids in regions of the genome called CpG islands, which contain highly frequent repetitions of cytosine and guanine and are typically found in gene promoters[143]. Histone modifications include acetylation, methylation, citrullination, phosphorylation, and ubiquitination; each is distinguished by the addition of a

chemical structure to specific histone amino acids to alter their structure and thus their activity[144].

MicroRNAs (miRNAs) are a family of 21- to 25-nucleotide noncoding small RNAs, which play a key role in post-transcriptional gene silencing by targeting message RNAs (mRNAs), primarily at the 3' untranslated regions, and thus controlling the translation of mRNAs into proteins[145].

In MS, these epigenetic mechanisms have been involved in the processes of inflammation, demyelination, remyelination, and neurodegeneration in MS, thus explaining part of the heritability of MS by altering the target gene expression.

1.6.5.1 Epigenetic changes associated with MS

1.6.5.1.1 Inflammation

Inflammation is an important mechanism of CNS damage in MS, and the proper transcriptional control of genes involved in the immune responses is mediated in part by epigenetic regulation. Kumagai and colleagues found that SHP-1, a negative regulator of proinflammatory signaling, had significantly greater promoter methylation in leukocytes in more than a third of MS subjects compared with healthy controls[146], potentially leading to decreased SHP-1 expression and increased leukocyte-mediated inflammation. A study of CD4+ T-cells from RRMS patients showed significantly greater demethylation of the *FOXP3* and *IL-17A* loci compared to healthy controls. They further showed that the demethylation of *FOXP3* can inhibit Th1 and Th2 cell differentiation while promoting Treg and Th17 cell lineage commitment, whereas the hypomethylation of *IL-17A* can lead to increased

differentiation towards the Th17 cell lineage[147]. In EAE, an animal model of CNS inflammation, brain inflammation was primarily mediated through the activities of Th17 cells, whereas Th1 cells were the primary causes of inflammation in the spinal cord. In addition to discrete levels of each T-cell type, the balance between Th17 and Treg cells can affect disease status; the balance between Th1 and Th2 can thus lead not only to injury but also repair in the CNS[148].

Most of the studies focusing on miRNAs in inflammation and MS are in agreement that miRNA dysregulation favors some T-cell differentiation and promotes disease progression. In CD4+ T-cells isolated from RRMS patients, miR-17-5p, which can target lipid kinases that regulate lymphocyte development, activation, and survival, is highly expressed in RRMS patients compared to controls[149]. Guerau-de Arellano and colleagues found increased expression of miR-128 and miR-27b microRNAs in the naïve T-cells of MS patients and increased expression of miR-340 in memory T-cells. The authors concluded that these altered expression levels lead to the inhibition of Th2 cell development and favor a pro-inflammatory Th1 response. miR-128, miR-27b, and miR-340 directly suppress the expression of B lymphoma Mo-MLV insertion region 1 homolog (BMI1), and miR-340 also suppresses the expression of interleukin-4 (IL4), resulting in an immune profile shift from Th2 to Th1[150]. Comparing the miRNA profile in Treg cells from 12 RRMS patients and 14 healthy controls, miR-25 and miR-106b were downregulated in cases compared to controls. These two miRNAs act on CDKN1A/p21 and BCL2L11/Bim and therefore modulate the TGF-β signaling pathway, which is involved in Treg cell differentiation and maturation[151]. By analyzing peripheral blood mononuclear cells (PBMCs) from 43 MS cases and 42 healthy controls, Du and colleagues found that Th17 differentiation was partially controlled by miR-326[152]. Moreover, miR-326 overexpression was

highly correlated with disease severity in patients with MS and EAE mice. They also found that miR-326 promoted Th17 differentiation by targeting and blocking the expression of Ets-1, whose function is the inhibition of the differentiation of naïve T-cells toward Th17[152].

1.6.5.1.2 Demyelination and remyelination

CNS inflammation may lead to demyelination and oligodendrocyte death, resulting in many of the clinical manifestations of MS. Early in the course of disease, these insults can be mild and reversible, with remyelination occurring thereafter[153]. Later in the disease course, however, the repair of this damage is increasingly incomplete or wholly unresolved, leading to increasing neurological damage and functional deficits.

Mastronardi and colleagues demonstrated that during the demyelination of white matter in MS patients, the promoter of peptidylarginine deiminase type-2 (PAD-2) was demethylated at cytosine residues in the CpG islands, leading to its overexpression in the brain. However, this change was only found in MS and was not found in other neurological diseases, including Alzheimer's disease, Parkinson's disease, and Huntington's disease[154]. PAD-2 can destabilize myelin basic protein (MBP) and the myelin multilayers, rendering myelin vulnerable to autoimmune attack[155]. Mastronardi and colleagues further revealed that in normal-appearing white matter from MS patients and EAE animal models, nucleosomal histone H3 can be citrullinated by another peptidylarginine deiminase, PAD-4.[156] In the early stages of MS, marked histone H3 deacetylation is found in oligodendrocytes. This increased histone H3 deacetylation correlates with the impaired differentiation of oligodendrocytes and may lead to impaired remyelination in MS patients[157].

Junker and colleagues established miRNA profiles from active and inactive MS lesions[158], finding that miR-214 and miR-23a were upregulated in both lesion types. During oligodendrocyte differentiation, both miR-214 and miR-23a are overexpressed, indicating that they may be involved in remyelination. They found that miR146a, which is primarily involved in the regulation of inflammation, was only upregulated in active brain lesions, whereas miR-219 and miR-338-5p were repressed only in inactive MS lesions. One of the target genes of miR-219 is ELOVL7, which maintains the integrity of myelin and axons in the CNS of adult mice[159]. miR-219 and miR-338-5p can also target the genes Sox6, Hes5, and Zfp238, inhibiting the differentiation of oligodendrocyte precursors and therefore affecting the remyelination process[160,161]. Another study showed that miR-155, miR-338, and miR-491 are upregulated in patients with a more progressive disease course. The overexpression of these miRNAs inhibited the translation of their target genes (aldo-keto reductase family members C1 and C2), which are involved in neurosteroid synthesis. In humans, neurosteroids have an important role in the regulation of myelination, neuroprotection, and the growth of axons and dendrites[162]. These researchers came to the conclusion that the shift of miRNA expression can significantly affect the regulation of demyelination and remyelination in MS.

1.6.5.1.3 Neurodegeneration

To date, there are no specific studies exploring the epigenetics of neurodegeneration associated with MS; however, we have interesting leads from other conditions. Chestnut and colleagues identified a link between DNA methylation and neuron death by using cultured NSC34 cells[163]. In cultured spinal cord neurons, the overexpression of Dnmt3a, which is responsible for DNA methylation, induced degeneration,

whereas the inhibition of Dnmt3a protected cultured neurons from apoptosis. Further studies in samples from cases with amyotrophic lateral sclerosis (ALS) found similar effects. These results indicate that DNA methylation might contribute to neurodegeneration[164] in MS patients. Because neurodegeneration develops concurrently with inflammatory demyelination, the epigenetic changes involved in inflammation and demyelination may also play important roles, although there is currently no direct evidence for this. The miRNAs that discussed above were summarized in Table 1.2

Table 1.2: Summary of the miRNAs discussed.

Diseases processes in MS	miRNAs
Inflammation	miR-17-5p,miR-128,miR-27b,miR-340,miR-25,miR-106b and miR-326
Demyelination and remyelination	miR-214,miR-23a,miR-146a,miR-219,miR-338-5p,miR-155,miR-338 and miR-491

1.6.5.2 Environmental risk factors for MS and epigenetic changes

Environmental risk factors can exert effects long before MS becomes clinically evident. Recent studies have revealed that the effects of these environmental risk factors on MS risk are mediated by epigenetic changes.

In EBV-transformed B-cells, Riley and colleagues found that EBV miRNAs cooperated with host miRNAs to target host mRNAs, particularly interacting with members of the miR-17-92 miRNA cluster, which induce lymphomas and inhibit apoptosis, and with miR-142-3p and miR-155, which are highly abundant immunomodulatory miRNAs[165]. In MS, miR-142-3p expression has been linked to immune tolerance[166], whereas miR-155 has been shown to associate with T-cell differentiation and promote CNS inflammation in EAE[167]. In addition, in EBV-

induced nasopharyngeal cancer[168], the enzymes DNMT1, DNMT3a, and DNMT3b increased the methylation of tumor suppressor gene promoters, thus reducing the expression of these genes.

Clinical trials in MS have demonstrated that the vitamin D has dose-dependent effects on the immune system[169]; higher vitamin D doses are associated with a greater reduction in T-cell proliferation[169,170]. Part of the effects of vitamin D on immune function may be via the induction of epigenetic changes that modulate the immune response. A single study has examined the effects of $1,25(OH)_2D_3$ on human IL-17A production using $CD4^+$ T-cells, and it provides some support for such a mechanism. This study showed that $1,25(OH)_2D_3$ directly repressed the transcription of the proinflammatory cytokine locus *IL-17* by the dissociation of histone acetylase activity in the *IL-17A* promoter and the recruitment of HDAC2 to the nuclear factor for activated T-cells (NFAT) binding sites, thus preventing its binding and the subsequent enhancement of *IL-17A* expression[171]. These results from one study alone emphasize the need for further work in this area.

By examining blood samples from adolescents whose mothers smoked during pregnancy, Toledo-Rodrigues and colleagues found that prenatal exposure to maternal cigarette smoking was associated with higher promoter methylation of brain-derived neurotrophic factor (BDNF), which supports the growth and differentiation of new neurons[172]. Maccani and colleagues found that three miRNAs, miR-16, miR-21, and mi4-146a, were downregulated as a result of maternal cigarette smoking during pregnancy[173]. Thus, cigarette smoking may modulate DNA methylation and miRNA activity in MS. However, there is currently no published research on the effects of such changes in MS.

1.6.5.3 Interaction of multiple environmental risk factors and epigenetic changes

Environmental risk factors and epigenetic changes do not occur through a single mechanism but through a complex interaction of several factors; they can interact and act synergistically. Epidemiological studies suggest that the risk for developing MS is much higher in the presence of multiple environmental risk factors. For example, the association between higher anti-EBNA IgG titers and increased MS risk is approximately two-fold greater among ever smokers compared to never smokers in some but not all studies[174,175]. Research suggests that miRNAs can regulate DNA methylation and histone modification[176-178]. In turn, these modifications can also regulate the expression of miRNAs[179]. As discussed earlier, each environmental risk factor can act via epigenetic effects to modulate MS. It is possible that the epigenetic changes caused by smoking influence the epigenetic changes caused by EBV or vitamin D deficiency, and vice-versa. Environmental risk factors and epigenetic changes associate with each other through a complex network, gradually accumulating effects and possibly amplifying the risk for MS. These networks and interactions may be highly complicated and act in different directions depending on the levels of exposure. Consequently, the study of gene-environment interactions and the interplay of epigenetics require careful studies of large MS cohorts in which all relevant confounders, both genetic and environmental, are measured and controlled for in analyses.

1.7 Summary

It is now well established that both environmental and genetic factors contribute to and interact in the development of MS. The well-established environmental factors

include infection with EBV and HHV-6, vitamin D deficiency, and cigarette smoking. Despite the success of genetic studies (linkage analysis and GWAS studies) in identifying the role of the MHC region and 110 non-MHC regions in the development of MS, the currently described causal genetic variants do not explain the majority of the heritability of MS, resulting in "missing heritability". Hypotheses to explain the missing heritability include lower frequency variants and rare variants, gene-gene interactions, gene-environmental interactions, and epigenetics. However, significant challenges exist in discovering the drivers of the missing heritability. For the lower frequency variants and rare variants hypothesis, large exome array studies in MS have yielded to date only a small number of potential rare variant targets that do not significantly add to the heritability of MS. For gene-gene interaction, multiple testing issues have limited its use at the GWAS scale. For gene-environmental interaction, the use of prospective studies assessing genetic and environmental factors in predicting the MS clinical course typically requires many years of follow up and the concurrent assessment of these factors in a systematic way. Unlike the GWAS that can be conducted using blood samples, epigenetic changes are cell-specific, and these changes are difficult to monitor in brain tissue.

Despite these challenges, we proposed three ways in this thesis to address some of the missing heritability. First, combine both environmental and genetic factors to predict MS using a cross-sectional study design and large sample size. For example, although most studies have shown that higher levels of antibodies against the EBV virus (anti-EBNA-1 IgG titer) are associated with higher MS risk, few studies have tried to determine which genes are associated with EBNA-1 titers and how these genes contribute to MS risk.

Second, use a prospective cohort study design with participants followed for a period of time to identify the single gene, gene-gene interactions, and gene-environmental interactions in predicting the clinical course.

Third, examine the functional variation within MS-related miRNA genes and determine how this variation can predispose one to MS.

1.8 Structure of the thesis

This thesis covers a variety of topics from my PhD research, but it mainly focuses on the three aspects discussed above: genetic susceptibility for MS (Chapter 2), genetic determinants of clinical course (Chapters 3-4), and functional variation in miRNA genes creating a predisposition to MS (Chapter 5).

In Chapter 2, I hypothesize that an altered immune response to EBV infection—manifested by increased anti-EBNA-1 IgG titers or a potentiated inflammatory response to EBV infection—may be influenced by genetic factors and that these genetic variants can also create a predisposition to MS. To test this hypothesis, I performed a genome-wide association study of anti-EBNA-1 IgG titers in 3,599 individuals from an unselected twin family cohort followed by a meta-analysis with data from an independent EBNA-1 GWAS. I then examined the shared polygenic risk between the EBNA-1 GWAS (effective sample size (N_{eff}) = 5,555) and a large MS GWAS (N_{eff} = 15,231).

In Chapter 3, I hypothesize that SNPs within genes for relevant cytokines and their receptors modulate the previously demonstrated associations of PBMC-produced cytokines, TNF-α and IFN-γ, with relapse, thus providing additional information

about these cytokine effects and the roles of these genes in MS. I tested this hypothesis using The Southern Tasmanian Multiple Sclerosis Longitudinal Study (MSL), which followed a cohort of 203 persons with clinically defined MS living in southern Tasmania, Australia from 2002 to 2005.

In Chapter 4, I hypothesize that variation in myelin basic protein altered the clinical course (conversion to MS after a first demyelinating event, relapse rate, and annualized change of disability) of MS. I tested this hypothesis using the Ausimmune/AusLong Longitudinal Study—an ongoing > 10-year prospective cohort study following a population of 282 patients initially identified with a first demyelinating event (FDE) suggestive of, but not diagnostic for, MS.

In Chapter 5, I hypothesize that polymorphisms in the *miR-146a* gene predict the MS clinical course (conversion to MS after a first demyelinating event, relapse rate, and annualized change of disability) of MS. We tested this hypothesis using the Ausimmune/AusLong Longitudinal Study as above.

Finally, in Chapter 6, I have summarized the main findings of my thesis and suggest directions for future work.

1.9 References

1. Ciccarelli O, Barkhof F, Bodini B, et al. Pathogenesis of multiple sclerosis: insights from molecular and metabolic imaging. *Lancet Neurol.* 2014;13(8):807-822.

2. Compston A, Coles A. Multiple sclerosis. *Lancet.* 2008;372(9648):1502-1517.

3. Charcot JM. Histologie de la sclérose en plaques [French]. *Gazette des Hopitaux.* 1868;41:554-555.

4. Dawson J. The histology of disseminated sclerosis. *Trans. Royal Soc. Edin.* 1916;50:517-740.

5. Rivers TM, Sprunt DH, Berry GP. Observations on Attempts to Produce Acute Disseminated Encephalomyelitis in Monkeys. *J. Exp. Med.* 1933;58(1):39-53.

6. Douglas M, Nigel C, Charles L. *Multiple Sclerosis.* Edinburgh1955.

7. The Atlas of MS 2013. http://www.msif.org/about-us/advocacy/atlas/atlas-of-ms/. Accessed 23 Jan, 2016.

8. Compston. A, Confavreux. C, McDonald. I, et al. *McAlpine's Multiple Sclerosis (Fourth Edition).* Elsevier Inc; 2006.

9. Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). *Neurology.* 1983;33(11):1444-1452.

10. Lublin FD, Reingold SC. Defining the clinical course of multiple sclerosis: results of an international survey. National Multiple Sclerosis Society (USA) Advisory Committee on Clinical Trials of New Agents in Multiple Sclerosis. *Neurology.* 1996;46(4):907-911.

11. Lublin FD, Reingold SC, Cohen JA, et al. Defining the clinical course of multiple sclerosis: the 2013 revisions. *Neurology.* 2014;83(3):278-286.

12. McDonald WI, Compston A, Edan G, et al. Recommended diagnostic criteria for multiple sclerosis: guidelines from the International Panel on the diagnosis of multiple sclerosis. *Ann Neurol.* 2001;50(1):121-127.

13. Polman CH, Reingold SC, Edan G, et al. Diagnostic criteria for multiple sclerosis: 2005 revisions to the "McDonald Criteria". *Ann. Neurol.* 2005;58(6):840-846.

14. Polman CH, Reingold SC, Banwell B, et al. Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. *Ann Neurol.* 2011;69(2):292-302.

15. Andreadou E. *Neuroimaging in Multiple Sclerosis, Neuroimaging - Clinical Applications.* 2012.

16. International Multiple Sclerosis Genetics C, Wellcome Trust Case Control C, Sawcer S, et al. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature.* 2011;476(7359):214-219.

17. International Multiple Sclerosis Genetics C, Beecham AH, Patsopoulos NA, et al. Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. *Nat. Genet.* 2013;45(11):1353-1360.

18. Henderson AP, Barnett MH, Parratt JD, Prineas JW. Multiple sclerosis: distribution of inflammatory cells in newly forming lesions. *Ann. Neurol.* 2009;66(6):739-753.

19. Buck D, Hemmer B. Treatment of multiple sclerosis: current concepts and future perspectives. *J. Neurol.* 2011;258(10):1747-1762.

20. Hemmer B, Kerschensteiner M, Korn T. Role of the innate and adaptive immune responses in the course of multiple sclerosis. *Lancet Neurol.* 2015;14(4):406-419.

21. Geissmann F, Gordon S, Hume DA, Mowat AM, Randolph GJ. Unravelling mononuclear phagocyte heterogeneity. *Nat. Rev. Immunol.* 2010;10(6):453-460.

22. Murray PJ, Wynn TA. Protective and pathogenic functions of macrophage subsets. *Nat. Rev. Immunol.* 2011;11(11):723-737.

23. Ransohoff RM, Cardona AE. The myeloid cells of the central nervous system parenchyma. *Nature.* 2010;468(7321):253-262.

24. Kettenmann H, Kirchhoff F, Verkhratsky A. Microglia: new roles for the synaptic stripper. *Neuron.* 2013;77(1):10-18.

25. Germain RN. T-cell development and the CD4-CD8 lineage decision. *Nat. Rev. Immunol.* 2002;2(5):309-322.

26. Fujinami RS, von Herrath MG, Christen U, Whitton JL. Molecular mimicry, bystander activation, or viral persistence: infections and autoimmune disease. *Clin. Microbiol. Rev.* 2006;19(1):80-94.

27. Gutcher I, Becher B. APC-derived cytokines and T cell polarization in autoimmune inflammation. *J. Clin. Invest.* 2007;117(5):1119-1127.

28. Burrell BE, Nakayama Y, Xu J, Brinkman CC, Bromberg JS. Regulatory T cell induction, migration, and function in transplantation. *J. Immunol.* 2012;189(10):4705-4711.

29. Maddur MS, Miossec P, Kaveri SV, Bayry J. Th17 cells: biology, pathogenesis of autoimmune and inflammatory diseases, and therapeutic strategies. *Am. J. Pathol.* 2012;181(1):8-18.

30. Nielsen HM, Tost J. Epigenetic changes in inflammatory and autoimmune diseases. *Subcell. Biochem.* 2013;61:455-478.

31. Ouyang W, Kolls JK, Zheng Y. The biological functions of T helper 17 cell effector cytokines in inflammation. *Immunity.* 2008;28(4):454-467.

32. Babbe H, Roers A, Waisman A, et al. Clonal expansions of CD8(+) T cells dominate the T cell infiltrate in active multiple sclerosis lesions as shown by micromanipulation and single cell polymerase chain reaction. *J. Exp. Med.* 2000;192(3):393-404.

33. Jacobsen M, Cepok S, Quak E, et al. Oligoclonal expansion of memory CD8+ T cells in cerebrospinal fluid from multiple sclerosis patients. *Brain.* 2002;125(Pt 3):538-550.

34. Frischer JM, Bramow S, Dal-Bianco A, et al. The relation between inflammation and neurodegeneration in multiple sclerosis brains. *Brain.* 2009;132(Pt 5):1175-1189.

35. Stys PK, Zamponi GW, van Minnen J, Geurts JJ. Will the real multiple sclerosis please stand up? *Nat. Rev. Neurosci.* 2012;13(7):507-514.

36. Confavreux C, Vukusic S, Moreau T, Adeleine P. Relapses and progression of disability in multiple sclerosis. *N. Engl. J. Med.* 2000;343(20):1430-1438.

37. Leray E, Yaouanq J, Le Page E, et al. Evidence for a two-stage disability progression in multiple sclerosis. *Brain.* 2010;133(Pt 7):1900-1913.

38. Ascherio A, Munger KL. Environmental risk factors for multiple sclerosis. Part II: Noninfectious factors. *Ann. Neurol.* 2007;61(6):504-513.

39. Ascherio A, Munger KL. Environmental risk factors for multiple sclerosis. Part I: the role of infection. *Ann. Neurol.* 2007;61(4):288-299.

40. Ebers GC. Environmental factors and multiple sclerosis. *Lancet Neurol.* 2008;7(3):268-277.

41. Amon W, Farrell PJ. Reactivation of Epstein-Barr virus from latency. *Rev. Med. Virol.* 2005;15(3):149-156.

42. Odumade OA, Hogquist KA, Balfour HH, Jr. Progress and problems in understanding and managing primary Epstein-Barr virus infections. *Clin. Microbiol. Rev.* 2011;24(1):193-209.

43. Ascherio A, Munger KL, Lunemann JD. The initiation and prevention of multiple sclerosis. *Nat. Rev. Neurol.* 2012;8(11):602-612.

44. Fraser KB, Haire M, Millar JH, McCrea S. Increased tendency to spontaneous in-vitro lymphocyte transformation in clinically active multiple sclerosis. *Lancet.* 1979;2(8145):175-176.

45. Levin LI, Munger KL, Rubertone MV, et al. Temporal relationship between elevation of Epstein-Barr virus antibody titers and initial onset of neurological

symptoms in multiple sclerosis. *JAMA: the journal of the American Medical Association.* 2005;293(20):2496-2500.

46. Levin LI, Munger KL, O'Reilly EJ, Falk KI, Ascherio A. Primary infection with the Epstein-Barr virus and risk of multiple sclerosis. *Ann Neurol.* 2010;67(6):824-830.

47. Belbasis L, Bellou V, Evangelou E, Ioannidis JP, Tzoulaki I. Environmental risk factors and multiple sclerosis: an umbrella review of systematic reviews and meta-analyses. *Lancet Neurol.* 2015;14(3):263-273.

48. Lunemann JD, Edwards N, Muraro PA, et al. Increased frequency and broadened specificity of latent EBV nuclear antigen-1-specific T cells in multiple sclerosis. *Brain.* 2006;129(Pt 6):1493-1506.

49. Lang HL, Jacobsen H, Ikemizu S, et al. A functional and structural basis for TCR cross-reactivity in multiple sclerosis. *Nat. Immunol.* 2002;3(10):940-943.

50. Serafini B, Rosicarelli B, Franciotta D, et al. Dysregulated Epstein-Barr virus infection in the multiple sclerosis brain. *J. Exp. Med.* 2007;204(12):2899-2912.

51. van Noort JM, Bajramovic JJ, Plomp AC, van Stipdonk MJ. Mistaken self, a novel model that links microbial infections with myelin-directed autoimmunity in multiple sclerosis. *J. Neuroimmunol.* 2000;105(1):46-57.

52. Pender MP. Infection of autoreactive B lymphocytes with EBV, causing chronic autoimmune diseases. *Trends Immunol.* 2003;24(11):584-588.

53. Salahuddin SZ, Ablashi DV, Markham PD, et al. Isolation of a new virus, HBLV, in patients with lymphoproliferative disorders. *Science.* 1986;234(4776):596-601.

54. Dominguez G, Dambaugh TR, Stamey FR, Dewhurst S, Inoue N, Pellett PE. Human herpesvirus 6B genome sequence: coding content and comparison with human herpesvirus 6A. *J. Virol.* 1999;73(10):8040-8052.

55. Gompels UA, Nicholas J, Lawrence G, et al. The DNA sequence of human herpesvirus-6: structure, coding content, and genome evolution. *Virology.* 1995;209(1):29-51.

56. Arbuckle JH, Medveczky PG. The molecular biology of human herpesvirus-6 latency and telomere integration. *Microbes Infect.* 2011;13(8-9):731-741.

57. Rotola A, Ravaioli T, Gonelli A, Dewhurst S, Cassai E, Di Luca D. U94 of human herpesvirus 6 is expressed in latently infected peripheral blood mononuclear cells and blocks viral gene expression in transformed

lymphocytes in culture. *Proc. Natl. Acad. Sci. U. S. A.* 1998;95(23):13911-13916.

58. Di Luca D, Mirandola P, Ravaioli T, et al. Human herpesviruses 6 and 7 in salivary glands and shedding in saliva of healthy and human immunodeficiency virus positive individuals. *J. Med. Virol.* 1995;45(4):462-468.

59. Challoner PB, Smith KT, Parker JD, et al. Plaque-associated expression of human herpesvirus 6 in multiple sclerosis. *Proc. Natl. Acad. Sci. U. S. A.* 1995;92(16):7440-7444.

60. Friedman JE, Lyons MJ, Cu G, et al. The association of the human herpesvirus-6 and MS. *Mult. Scler.* 1999;5(5):355-362.

61. Enbom M, Wang FZ, Fredrikson S, Martin C, Dahl H, Linde A. Similar humoral and cellular immunological reactivities to human herpesvirus 6 in patients with multiple sclerosis and controls. *Clin. Diagn. Lab. Immunol.* 1999;6(4):545-549.

62. Chapenko S, Millers A, Nora Z, Logina I, Kukaine R, Murovska M. Correlation between HHV-6 reactivation and multiple sclerosis disease activity. *J. Med. Virol.* 2003;69(1):111-117.

63. Rotola A, Cassai E, Tola MR, Granieri E, Di Luca D. Human herpesvirus 6 is latent in peripheral blood of patients with relapsing-remitting multiple sclerosis. *J. Neurol. Neurosurg. Psychiatry.* 1999;67(4):529-531.

64. Rotola A, Merlotti I, Caniatti L, et al. Human herpesvirus 6 infects the central nervous system of multiple sclerosis patients in the early stages of the disease. *Mult. Scler.* 2004;10(4):348-354.

65. Ortega-Madueno I, Garcia-Montojo M, Dominguez-Mozo MI, et al. Anti-human herpesvirus 6A/B IgG correlates with relapses and progression in multiple sclerosis. *PLoS One.* 2014;9(8):e104836.

66. Alvarez-Lafuente R, Garcia-Montojo M, De las Heras V, Bartolome M, Arroyo R. Clinical parameters and HHV-6 active replication in relapsing-remitting multiple sclerosis patients. *J Clin Virol.* 2006;37 Suppl 1:S24-26.

67. Simpson S, Jr., Taylor B, Dwyer DE, et al. Anti-HHV-6 IgG titer significantly predicts subsequent relapse risk in multiple sclerosis. *Mult. Scler.* 2012;18(6):799-806.

68. Tejada-Simon MV, Zang YC, Hong J, Rivera VM, Zhang JZ. Cross-reactivity with myelin basic protein and human herpesvirus-6 in multiple sclerosis. *Ann Neurol.* 2003;53(2):189-197.

69. van der Mei IA, Ponsonby AL, Blizzard L, Dwyer T. Regional variation in multiple sclerosis prevalence in Australia and its association with ambient ultraviolet radiation. *Neuroepidemiology.* 2001;20(3):168-174.

70. Marrie RA. Environmental risk factors in multiple sclerosis aetiology. *Lancet Neurol.* 2004;3(12):709-718.

71. Bischoff-Ferrari HA, Giovannucci E, Willett WC, Dietrich T, Dawson-Hughes B. Estimation of optimal serum concentrations of 25-hydroxyvitamin D for multiple health outcomes. *Am. J. Clin. Nutr.* 2006;84(1):18-28.

72. Munger KL, Levin LI, Hollis BW, Howard NS, Ascherio A. Serum 25-hydroxyvitamin D levels and risk of multiple sclerosis. *JAMA.* 2006;296(23):2832-2838.

73. Hernan MA, Olek MJ, Ascherio A. Cigarette smoking and incidence of multiple sclerosis. *Am. J. Epidemiol.* 2001;154(1):69-74.

74. Hedstrom AK, Baarnhielm M, Olsson T, Alfredsson L. Tobacco smoking, but not Swedish snuff use, increases the risk of multiple sclerosis. *Neurology.* 2009;73(9):696-701.

75. Chen JL, Wei L, Bereczki D, et al. Nicotine raises the influx of permeable solutes across the rat blood-brain barrier with little or no capillary recruitment. *J. Cereb. Blood Flow Metab.* 1995;15(4):687-698.

76. Costenbader KH, Karlson EW. Cigarette smoking and autoimmune disease: what can we learn from epidemiology? *Lupus.* 2006;15(11):737-745.

77. Robertson NP, Clayton D, Fraser M, Deans J, Compston DA. Clinical concordance in sibling pairs with multiple sclerosis. *Neurology.* 1996;47(2):347-352.

78. Carton H, Vlietinck R, Debruyne J, et al. Risks of multiple sclerosis in relatives of patients in Flanders, Belgium. *J. Neurol. Neurosurg. Psychiatry.* 1997;62(4):329-333.

79. Robertson NP, Fraser M, Deans J, Clayton D, Walker N, Compston DA. Age-adjusted recurrence risks for relatives of patients with multiple sclerosis. *Brain.* 1996;119 (Pt 2):449-455.

80. Willer CJ, Dyment DA, Risch NJ, Sadovnick AD, Ebers GC, Canadian Collaborative Study G. Twin concordance and sibling recurrence rates in multiple sclerosis. *Proc. Natl. Acad. Sci. U. S. A.* 2003;100(22):12877-12882.

81. Mumford CJ, Wood NW, Kellar-Wood H, Thorpe JW, Miller DH, Compston DA. The British Isles survey of multiple sclerosis in twins. *Neurology.* 1994;44(1):11-15.

82. Ebers GC, Sadovnick AD, Risch NJ. A genetic basis for familial aggregation in multiple sclerosis. Canadian Collaborative Study Group. *Nature.* 1995;377(6545):150-151.

83. Robertson NP, O'Riordan JI, Chataway J, et al. Offspring recurrence rates and clinical characteristics of conjugal multiple sclerosis. *Lancet.* 1997;349(9065):1587-1590.

84. Pulst SM. Genetic linkage analysis. *Arch. Neurol.* 1999;56(6):667-672.

85. Ott J, Wang J, Leal SM. Genetic linkage analysis in the age of whole-genome sequencing. *Nat Rev Genet.* 2015;16(5):275-284.

86. Sawcer S, Franklin RJ, Ban M. Multiple sclerosis genetics. *Lancet Neurol.* 2014;13(7):700-709.

87. Sawcer S, Ban M, Maranian M, et al. A high-density screen for linkage in multiple sclerosis. *Am. J. Hum. Genet.* 2005;77(3):454-467.

88. Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet.* 2005;6(2):95-108.

89. Risch N, Merikangas K. The future of genetic studies of complex human diseases. *Science.* 1996;273(5281):1516-1517.

90. Welter D, MacArthur J, Morales J, et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.* 2014;42(Database issue):D1001-1006.

91. Price AL, Zaitlen NA, Reich D, Patterson N. New approaches to population stratification in genome-wide association studies. *Nat Rev Genet.* 2010;11(7):459-463.

92. Devlin B, Roeder K. Genomic control for association studies. *Biometrics.* 1999;55(4):997-1004.

93. Pritchard JK, Rosenberg NA. Use of unlinked genetic markers to detect population stratification in association studies. *Am. J. Hum. Genet.* 1999;65(1):220-228.

94. Reich DE, Goldstein DB. Detecting association in a case-control study while correcting for population stratification. *Genet. Epidemiol.* 2001;20(1):4-16.

95. Voight BF, Pritchard JK. Confounding from cryptic relatedness in case-control association studies. *PLoS Genet.* 2005;1(3):e32.

96. Gao X, Becker LC, Becker DM, Starmer JD, Province MA. Avoiding the high Bonferroni penalty in genome-wide association studies. *Genet. Epidemiol.* 2010;34(1):100-105.

97. Hochberg Y, Benjamini Y. More powerful procedures for multiple significance testing. *Stat. Med.* 1990;9(7):811-818.

98. Sham PC, Purcell SM. Statistical power and significance testing in large-scale genetic studies. *Nat Rev Genet.* 2014;15(5):335-346.

99. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 2007;81(3):559-575.

100. International Multiple Sclerosis Genetics C, Hafler DA, Compston A, et al. Risk alleles for multiple sclerosis identified by a genomewide study. *N Engl J Med.* 2007;357(9):851-862.

101. Brynedal B, Wojcik J, Esposito F, et al. MGAT5 alters the severity of multiple sclerosis. *J. Neuroimmunol.* 2010;220(1-2):120-124.

102. Horton R, Wilming L, Rand V, et al. Gene map of the extended human MHC. *Nat Rev Genet.* 2004;5(12):889-899.

103. Miretti MM, Walsh EC, Ke X, et al. A high-resolution linkage-disequilibrium map of the human major histocompatibility complex and first generation of tag single-nucleotide polymorphisms. *Am. J. Hum. Genet.* 2005;76(4):634-646.

104. Patsopoulos NA, Barcellos LF, Hintzen RQ, et al. Fine-mapping the genetic association of the major histocompatibility complex in multiple sclerosis: HLA and non-HLA effects. *PLoS Genet.* 2013;9(11):e1003926.

105. Smith KJ, Pyrdol J, Gauthier L, Wiley DC, Wucherpfennig KW. Crystal structure of HLA-DR2 (DRA*0101, DRB1*1501) complexed with a peptide from human myelin basic protein. *J. Exp. Med.* 1998;188(8):1511-1520.

106. Oksenberg JR, Baranzini SE, Sawcer S, Hauser SL. The genetics of multiple sclerosis: SNPs to pathways to pathogenesis. *Nat Rev Genet.* 2008;9(7):516-526.

107. McLaren W, Pritchard B, Rios D, Chen Y, Flicek P, Cunningham F. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics.* 2010;26(16):2069-2070.

108. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nat Methods.* 2010;7(4):248-249.

109. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the functional effect of amino acid substitutions and indels. *PLoS One.* 2012;7(10):e46688.

110. Friese MA, Schattling B, Fugger L. Mechanisms of neurodegeneration and axonal dysfunction in multiple sclerosis. *Nat. Rev. Neurol.* 2014;10(4):225-238.

111. Baranzini SE, Wang J, Gibson RA, et al. Genome-wide association analysis of susceptibility and clinical phenotype in multiple sclerosis. *Hum. Mol. Genet.* 2009;18(4):767-778.

112. International Multiple Sclerosis Genetics C. Genome-wide association study of severity in multiple sclerosis. *Genes Immun.* 2011;12(8):615-625.

113. George MF, Briggs FB, Shao X, et al. Multiple sclerosis risk loci and disease severity in 7,125 individuals from 10 studies. *Neurol Genet.* 2016;2(4):e87.

114. Lill CM. Recent advances and future challenges in the genetics of multiple sclerosis. *Front. Neurol.* 2014;5:130.

115. Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases. *Nature.* 2009;461(7265):747-753.

116. Ramagopalan SV, Dyment DA, Cader MZ, et al. Rare variants in the CYP27B1 gene are associated with multiple sclerosis. *Ann. Neurol.* 2011;70(6):881-886.

117. Ban M, Caillier S, Mero IL, et al. No evidence of association between mutant alleles of the CYP27B1 gene and multiple sclerosis. *Ann. Neurol.* 2013;73(3):430-432.

118. Cortes A, Field J, Glazov EA, et al. Resequencing and fine-mapping of the chromosome 12q13-14 locus associated with multiple sclerosis refines the number of implicated genes. *Hum. Mol. Genet.* 2013;22(11):2283-2292.

119. Barizzone N, Pauwels I, Luciano B, et al. No evidence for a role of rare CYP27B1 functional variations in multiple sclerosis. *Ann. Neurol.* 2013;73(3):433-437.

120. Hunt KA, Mistry V, Bockett NA, et al. Negligible impact of rare autoimmune-locus coding-region variants on missing heritability. *Nature.* 2013;498(7453):232-235.

121. Genomes Project C, Abecasis GR, Altshuler D, et al. A map of human genome variation from population-scale sequencing. *Nature.* 2010;467(7319):1061-1073.

122. Wei WH, Hemani G, Haley CS. Detecting epistasis in human complex traits. *Nat Rev Genet.* 2014;15(11):722-733.

123. Cordell HJ. Detecting gene-gene interactions that underlie human diseases. *Nat Rev Genet.* 2009;10(6):392-404.

124. Emily M, Mailund T, Hein J, Schauser L, Schierup MH. Using biological networks to search for interacting loci in genome-wide association studies. *Eur. J. Hum. Genet.* 2009;17(10):1231-1240.

125. Lincoln MR, Ramagopalan SV, Chao MJ, et al. Epistasis among HLA-DRB1, HLA-DQA1, and HLA-DQB1 loci determines multiple sclerosis susceptibility. *Proc. Natl. Acad. Sci. U. S. A.* 2009;106(18):7542-7547.

126. Barcellos LF, Sawcer S, Ramsay PP, et al. Heterogeneity at the HLA-DRB1 locus and risk for multiple sclerosis. *Hum. Mol. Genet.* 2006;15(18):2813-2824.

127. Ramagopalan SV, Morris AP, Dyment DA, et al. The inheritance of resistance alleles in multiple sclerosis. *PLoS Genet.* 2007;3(9):1607-1613.

128. DeLuca GC, Ramagopalan SV, Herrera BM, et al. An extremes of outcome strategy provides evidence that multiple sclerosis severity is determined by alleles at the HLA-DRB1 locus. *Proc. Natl. Acad. Sci. U. S. A.* 2007;104(52):20896-20901.

129. International Multiple Sclerosis Genetics C, International IBDGC, International IBDGCI. Class II HLA interactions modulate genetic risk for multiple sclerosis. *Nat. Genet.* 2015.

130. Australia, New Zealand Multiple Sclerosis Genetics C. Genome-wide association study identifies new multiple sclerosis susceptibility loci on chromosomes 12 and 20. *Nat. Genet.* 2009;41(7):824-828.

131. Johnson BA, Wang J, Taylor EM, et al. Multiple sclerosis susceptibility alleles in African Americans. *Genes Immun.* 2010;11(4):343-350.

132. Bush WS, McCauley JL, DeJager PL, et al. A knowledge-driven interaction analysis reveals potential neurodegenerative mechanism of multiple sclerosis susceptibility. *Genes Immun.* 2011;12(5):335-340.

133. Thomas D. Gene--environment-wide association studies: emerging approaches. *Nat Rev Genet.* 2010;11(4):259-272.

134. Ellis JA, Kemp AS, Ponsonby AL. Gene-environment interaction in autoimmune disease. *Expert Rev. Mol. Med.* 2014;16:e4.

135. Manolio TA, Bailey-Wilson JE, Collins FS. Genes, environment and the value of prospective cohort studies. *Nat Rev Genet.* 2006;7(10):812-820.

136. van der Mei IA, Ponsonby AL, Taylor BV, et al. Human leukocyte antigen-DR15, low infant sibling exposure and multiple sclerosis: gene-environment interaction. *Ann. Neurol.* 2010;67(2):261-265.

137. van der Mei I, Lucas RM, Taylor BV, et al. Population attributable fractions and joint effects of key risk factors for multiple sclerosis. *Mult. Scler.* 2015.

138. Sackett DL. Bias in analytic research. *J. Chronic Dis.* 1979;32(1-2):51-63.

139. Lin R, Taylor BV, Simpson S, Jr., et al. Association between multiple sclerosis risk-associated SNPs and relapse and disability--a prospective cohort study. *Mult. Scler.* 2014;20(3):313-321.

140. Lin R, Taylor BV, Simpson S, Jr., et al. Novel modulating effects of PKC family genes on the relationship between serum vitamin D and relapse in multiple sclerosis. *J. Neurol. Neurosurg. Psychiatry.* 2014;85(4):399-404.

141. Jaenisch R, Bird A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat. Genet.* 2003;33 Suppl:245-254.

142. Koch MW, Metz LM, Kovalchuk O. Epigenetic changes in patients with multiple sclerosis. *Nature reviews. Neurology.* 2013;9(1):35-43.

143. Bird A. DNA methylation patterns and epigenetic memory. *Genes Dev.* 2002;16(1):6-21.

144. Dieker J, Muller S. Epigenetic histone code and autoimmunity. *Clinical reviews in allergy & immunology.* 2010;39(1):78-84.

145. Baek D, Villen J, Shin C, Camargo FD, Gygi SP, Bartel DP. The impact of microRNAs on protein output. *Nature.* 2008;455(7209):64-71.

146. Kumagai C, Kalman B, Middleton FA, Vyshkina T, Massa PT. Increased promoter methylation of the immune regulatory gene SHP-1 in leukocytes of multiple sclerosis subjects. *J. Neuroimmunol.* 2012;246(1-2):51-57.

147. Janson PC, Linton LB, Bergman EA, et al. Profiling of CD4+ T cells with epigenetic immune lineage analysis. *J. Immunol.* 2011;186(1):92-102.

148. Jadidi-Niaragh F, Mirshafiey A. Th17 cell, the new player of neuroinflammatory process in multiple sclerosis. *Scand. J. Immunol.* 2011;74(1):1-13.

149. Lindberg RL, Hoffmann F, Mehling M, Kuhle J, Kappos L. Altered expression of miR-17-5p in CD4+ lymphocytes of relapsing-remitting multiple sclerosis patients. *Eur. J. Immunol.* 2010;40(3):888-898.

150. Guerau-de-Arellano M, Smith KM, Godlewski J, et al. Micro-RNA dysregulation in multiple sclerosis favours pro-inflammatory T-cell-mediated autoimmunity. *Brain: a journal of neurology.* 2011;134(Pt 12):3578-3589.

151. De Santis G, Ferracin M, Biondani A, et al. Altered miRNA expression in T regulatory cells in course of multiple sclerosis. *J. Neuroimmunol.* 2010;226(1-2):165-171.

152. Du C, Liu C, Kang J, et al. MicroRNA miR-326 regulates TH-17 differentiation and is associated with the pathogenesis of multiple sclerosis. *Nat. Immunol.* 2009;10(12):1252-1259.

153. Bruck W. The pathology of multiple sclerosis is the result of focal inflammatory demyelination with axonal damage. *J. Neurol.* 2005;252 Suppl 5:v3-9.

154. Mastronardi FG, Noor A, Wood DD, Paton T, Moscarello MA. Peptidyl argininedeiminase 2 CpG island in multiple sclerosis white matter is hypomethylated. *J. Neurosci. Res.* 2007;85(9):2006-2016.

155. Kim JK, Mastronardi FG, Wood DD, Lubman DM, Zand R, Moscarello MA. Multiple sclerosis: an important role for post-translational modifications of myelin basic protein in pathogenesis. *Molecular & cellular proteomics: MCP.* 2003;2(7):453-462.

156. Mastronardi FG, Wood DD, Mei J, et al. Increased citrullination of histone H3 in multiple sclerosis brain and animal models of demyelination: a role for tumor necrosis factor-induced peptidylarginine deiminase 4 translocation. *The Journal of neuroscience: the official journal of the Society for Neuroscience.* 2006;26(44):11387-11396.

157. Pedre X, Mastronardi F, Bruck W, Lopez-Rodas G, Kuhlmann T, Casaccia P. Changed histone acetylation patterns in normal-appearing white matter and early multiple sclerosis lesions. *The Journal of neuroscience: the official journal of the Society for Neuroscience.* 2011;31(9):3435-3445.

158. Junker A, Krumbholz M, Eisele S, et al. MicroRNA profiling of multiple sclerosis lesions identifies modulators of the regulatory protein CD47. *Brain.* 2009;132(Pt 12):3342-3352.

159. Shin D, Shin JY, McManus MT, Ptacek LJ, Fu YH. Dicer ablation in oligodendrocytes provokes neuronal impairment in mice. *Ann. Neurol.* 2009;66(6):843-857.

160. Dugas JC, Cuellar TL, Scholze A, et al. Dicer1 and miR-219 Are required for normal oligodendrocyte differentiation and myelination. *Neuron.* 2010;65(5):597-611.

161. Zhao X, He X, Han X, et al. MicroRNA-mediated control of oligodendrocyte differentiation. *Neuron.* 2010;65(5):612-626.

162. Noorbakhsh F, Ellestad KK, Maingat F, et al. Impaired neurosteroid synthesis in multiple sclerosis. *Brain.* 2011;134(Pt 9):2703-2721.

163. Chestnut BA, Chang Q, Price A, Lesuisse C, Wong M, Martin LJ. Epigenetic regulation of motor neuron cell death through DNA methylation. *The Journal of neuroscience: the official journal of the Society for Neuroscience.* 2011;31(46):16619-16636.

164. De Jager PL, Srivastava G, Lunnon K, et al. Alzheimer's disease: early alterations in brain DNA methylation at ANK1, BIN1, RHBDF2 and other loci. *Nat. Neurosci.* 2014;17(9):1156-1163.

165. Riley KJ, Rabinowitz GS, Yario TA, Luna JM, Darnell RB, Steitz JA. EBV and human microRNAs co-target oncogenic and apoptotic viral and human genes during latency. *The EMBO journal.* 2012;31(9):2207-2221.

166. Waschbisch A, Atiya M, Linker RA, Potapov S, Schwab S, Derfuss T. Glatiramer acetate treatment normalizes deregulated microRNA expression in relapsing remitting multiple sclerosis. *PLoS One.* 2011;6(9):e24604.

167. Murugaiyan G, Beynon V, Mittal A, Joller N, Weiner HL. Silencing microRNA-155 ameliorates experimental autoimmune encephalomyelitis. *Journal of immunology.* 2011;187(5):2213-2221.

168. Kwong J, Lo KW, To KF, Teo PM, Johnson PJ, Huang DP. Promoter hypermethylation of multiple genes in nasopharyngeal carcinoma. *Clinical cancer research: an official journal of the American Association for Cancer Research.* 2002;8(1):131-137.

169. Ascherio A, Munger KL, Lunemann JD. The initiation and prevention of multiple sclerosis. *Nat Rev Neurol.* 2012;8(11):602-612.

170. Smolders J, Peelen E, Thewissen M, et al. Safety and T cell modulating effects of high dose vitamin D3 supplementation in multiple sclerosis. *PLoS One.* 2010;5(12):e15235.

171. Joshi S, Pantalena LC, Liu XK, et al. 1,25-dihydroxyvitamin D(3) ameliorates Th17 autoimmunity via transcriptional modulation of interleukin-17A. *Mol Cell Biol.* 2011;31(17):3653-3669.

172. Toledo-Rodriguez M, Lotfipour S, Leonard G, et al. Maternal smoking during pregnancy is associated with epigenetic modifications of the brain-derived neurotrophic factor-6 exon in adolescent offspring. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* 2010;153B(7):1350-1354.

173. Maccani MA, Avissar-Whiting M, Banister CE, McGonnigal B, Padbury JF, Marsit CJ. Maternal cigarette smoking during pregnancy is associated with downregulation of miR-16, miR-21, and miR-146a in the placenta. *Epigenetics.* 2010;5(7):583-589.

174. Simon KC, van der Mei IA, Munger KL, et al. Combined effects of smoking, anti-EBNA antibodies, and HLA-DRB1*1501 on multiple sclerosis risk. *Neurology.* 2010;74(17):1365-1371.

175. Sundqvist E, Sundstrom P, Linden M, et al. Epstein-Barr virus and multiple sclerosis: interaction with HLA. *Genes and immunity.* 2012;13(1):14-20.

176. Ngo-Yin Fan D, Ho-Ching Tsang F, Hoi-Kam Tam A, et al. Histone lysine methyltransferase, suppressor of variegation 3-9 homolog 1, promotes hepatocellular carcinoma progression and is negatively regulated by microRNA-125b. *Hepatology.* 2013;57(2):637-647.

177. Zhao S, Wang Y, Liang Y, et al. MicroRNA-126 regulates DNA methylation in CD4+ T cells and contributes to systemic lupus erythematosus by targeting DNA methyltransferase 1. *Arthritis Rheum.* 2011;63(5):1376-1386.

178. Huang J, Wang Y, Guo Y, Sun S. Down-regulated microRNA-152 induces aberrant DNA methylation in hepatitis B virus-related hepatocellular carcinoma by targeting DNA methyltransferase 1. *Hepatology.* 2010;52(1):60-70.

179. Scott GK, Mattie MD, Berger CE, Benz SC, Benz CC. Rapid alteration of microRNA levels by histone deacetylase inhibition. *Cancer Res.* 2006;66(3):1277-1281.

Appendix 1.A: Publication as part of Chapter 1: “The potential role of epigenetic modifications in the heritability of MS”

Zhou Y, Simpson S, Jr., Holloway AF, Charlesworth J, van der Mei I, Taylor BV. The potential role of epigenetic modifications in the heritability of multiple sclerosis. Mult Scler. 2014;20(2):135-40.

Chapter 2: Genetic loci for Epstein-Barr Virus nuclear antigen-1 are associated with risk of multiple sclerosis

2.1 Preface

The Epstein-Barr virus (EBV) has been extensively studied as an environmental causative factor in MS. However, which genes are associated with the level of immune response to EBV and how these genes contribute to the development of MS remain largely unknown. Thus, we performed a genome-wide association study (GWAS) in a large twin family cohort recruited in Queensland, Australia to study the genetic drivers of elevated titers of Epstein Barr Nuclear Antigen 1 (EBNA-1). EBNA-1 titers have been strongly associated with MS risk, and conversion to EBNA-1 positivity is directly correlated with EBV infection. We conducted a meta-analysis combining these results with the Mexican-American family GWAS that examined the same question. We examined the shared polygenic risk utilizing our EBNA-1 GWAS meta-analysis results and the results from a large, well-validated MS GWAS meta-analysis to determine the genetic architecture of anti-EBNA-1 immune responses and whether these loci were over-represented in MS cases compared to healthy controls, thus potentially representing novel MS susceptibility loci. This chapter was published in Multiple Sclerosis Journal in 2016[1] (Appendix 3.A).

2.2 Introduction

Epstein-Barr virus (EBV) is a double-stranded DNA human herpesvirus also known as HHV-4, which infects more than 90% of the world's adults and persists as a

lifelong latent infection[2]. It primarily infects B-lymphocytes and permanently transforms them into latently infected lymphoblastoid cell lines, which produce a group of viral proteins, including EBV nuclear antigens (EBNAs) and latent membrane proteins[3]. Infection with EBV has been associated with the development of autoimmune diseases including multiple sclerosis (MS), systemic lupus erythematous and rheumatoid arthritis[4]. Despite the high prevalence of EBV infection—only a small number of infected individuals will develop one of these diseases[5].

EBV nuclear antigen-1 (EBNA-1) may act as an important viral antigen in MS pathogenesis[6,7]. Recent work has suggested that HLA-dependent and independent immune responses to EBNA-1 are important drivers of MS pathogenesis and may modulate effects on MS risk of protective and deleterious HLA antigens[8]. Additionally, EBNA-1-specific T-cells that cross react with myelin have been demonstrated[9], and anti-EBNA-1 antibodies have been described in cerebrospinal fluid oligoclonal bands in MS cases[10]. Higher anti-EBNA-1 titers have been associated with a worse outcome in MS, both clinically[11] and radiologically[12].

A number of studies have observed an association between anti-EBNA-1 IgG titers and MS risk[7,13,14]. In a study of US military personnel who were anti-EBNA-1 IgG positive at study entry, the relative risk of MS over an average follow-up of five years was 36 times higher among those with high anti-EBNA titers compared with those with low titers. Importantly, in this study, in blood samples collected before the age of 20, the mean anti-EBNA-1 IgG titers were identical between participants who later developed MS and age and sex-matched controls that remained healthy. However, after the age of 20, the anti-EBNA-1 IgG titers in those that later developed MS increased four-fold, while that of controls remained constant[13]. Similarly, in ten cases

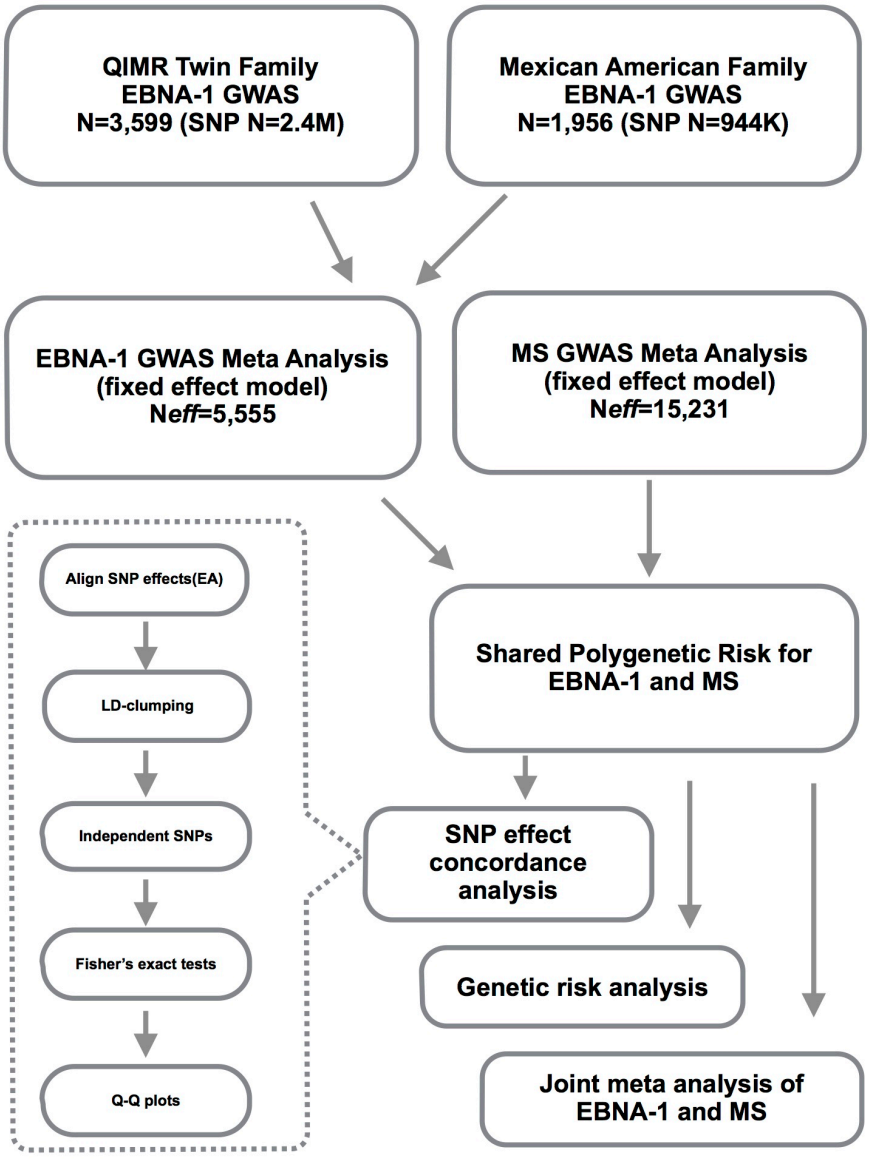
who were seronegative for anti-EBNA-1 IgG at their first blood sample who subsequently developed MS, all became seropositive generally within a few months of MS onset[15]. These findings suggest that the elevation in anti-EBNA-1 IgG titers may either directly predispose to the development of MS or may be a biomarker of immune dysregulation that results in an increased risk of MS[16].

Recent work in Mexican-American families[17] suggested that particular HLA variants are associated with elevated anti-EBNA-1 IgG titers. Thus, we hypothesized that an altered immune response to EBV infection—manifested by increased anti-EBNA-1 IgG titers or a potentiated inflammatory response to EBV infection—may be influenced by genetic factors, and that these genetic variants may also predispose to MS.

2.3 Materials and methods

A flow chart of the study design is shown in Figure 2.1.

Figure 2.1 Flow chart of the study design. EBNA-1: EBV nuclear antigen-1; GWAS: genome wide association analysis study; SNP: single nucleotide polymorphism; EA: Effect allele; LD: linkage disequilibrium; Q-Q plot: quantile-quantile (q-q) plot.



2.3.1 Genome-wide association analysis

2.3.1.1 QIMR Twin Families EBNA-1 GWAS (QTFEGWAS)

In total, 3,760 individuals from 1,020 families in the Queensland Institute of Medical Research (QIMR) Twin Family Cohort Study were studied. Details of the clinical protocol for blood collection and processing have been described elsewhere[18,19]. For a more detailed description of the sample, please see Supplementary Table 2.1. The ImmunoWELL EBNA IgG Test kit (GenBio, San Diego, CA) was used for detection of anti-EBNA-1 IgG in human serum (N = 3,760) at the Monash Antibody Technologies Facility, Monash University. The following values were used to determine the status for seropositive/seronegative according to the manufacturer's protocol: negative (<200 units/mL); equivocal (200-300 units/mL); and positive (>300 units/mL). To determine which data transformation should be used for analysis, the distribution of standardized residuals was examined by treating anti-EBNA-1 IgG titers as both continuous and categorical values (four groups: <250; 250-1,001; 1,001-2,000; >2,000 (units/mL)). The residuals were derived from generalized linear model (GLM) analysis by using box effect, row effect and column effect as fixed factors while also controlling for sex, age, age×sex, age^2 and age^2×sex. Because the residuals were normally distributed (Supplementary Figure 2.1), we utilized these modelled residuals as continuous outcomes and predictors in place of as-measured anti-EBNA-1 IgG.

Genome-wide genotyping was performed using the Illumina 610-Quad BeadChip and imputed to extend the genomic coverage to 2.4 million SNPs[20]. As part of quality control, individuals from families with pedigree errors and a call rate <0.95 were

excluded. SNPs with a call rate <0.95 or with a low imputation quality score (R^2 < 0.3), minor allele frequency <0.01 in the population, or significant deviation from Hardy-Weinberg equilibrium ($P_{HWE} < 1.0\times10^{-6}$), were excluded. In total, the data from 3,599 individuals from 992 families, and 2,428,106 SNPs passed quality control and were used for further analysis.

The association analysis of imputed dosage scores was performed using a family-based score test implemented in Merlin-offline[21] which corrects for relatedness of twins and family members. To detect whether there were independent effects at significant loci we carried out conditional analysis by including the most significant SNP as a covariate.

2.3.1.2 Mexican-American families EBNA-1 GWAS (MAFEGWAS)

We used summary results from the Mexican-American families EBNA-1 GWAS[17]. In brief, the SNPs were typed on several versions of Illumina's microarrays and quality controls were done according to standard requirements. A total of 944,565 genotyped SNPs from 1,956 individuals within 63 families were available for analysis. The anti-EBNA-1 quantitative antibody titer level was used for genome-wide joint linkage and association analysis in SOLAR[22].

2.3.2 GWAS meta-analysis

2.3.2.1 EBNA-1 GWAS Meta-Analysis (EGMA)

Summary results from the two EBNA-1 GWAS [QTFEGWAS (N = 3,599) and MAFEGWAS (N = 1,956)] were used to carry out a meta-analysis using an inverse

variance fixed-effect model in META[23]. Because we synthesized less than four studies, random effect modeling (and all tests and metrics of statistical heterogeneity) was inappropriate. T^2 (the between studies variance) is heavily underestimated, and this affects Cochran Q and I^2. The random effects model which uses T^2 to weight, hence suffers as well. To facilitate comparison with p-values for all 110 established non-HLA risk loci recently published by the International Multiple Sclerosis Genetics Consortium[24], we used the Direct Imputation of summary Statistics (DIST) program, to directly impute EGMA summary statistics based on the 1,000 genome project reference data. The direct imputation of summary statistics is achieved by employing a conditional expectation formula for multivariate normal variants and using the correlation structure from a relevant reference population. Details of the formula used please refer to paper[25]. Default and recommended parameters of DIST were utilized in this study.

2.3.2.2 MS GWAS Meta-Analysis (MGMA)

The MS GWAS meta-analysis was comprised of data from seven datasets of non-overlapping case and control subjects of European descent. A total of 5,545 cases and 12,153 controls with 2,529,394 SNPs were used for analysis. For a detailed description, see Patsopoulos et al[26].

2.3.2.3 Shared polygenic risk between EBNA-1 GWAS and MS GWAS

To determine whether the SNPs associated with anti-EBNA-1 IgG titers may also contribute to MS risk, we first matched the genome-wide significant SNPs ($P < 5\times10^{-8}$) in the EGMA with MGMA results. We also matched EGMA p-values for all 110

established non-HLA risk loci recently published [24]. We then carried out SNP effect concordant analysis and genetic risk score analysis to further investigate the shared polygenic risk between the EBNA-1 GWAS and the MS GWAS, including and excluding the extended HLA region (chr6: 25-35Mb).

2.3.2.4 SNP effect concordant analysis (SECA)

SECA was undertaken using the EGMA and MGMA, utilizing one study as a discovery sample while the other acted as a target sample, and vice-versa. Subsets of independent autosomal SNPs were extracted via linkage disequilibrium (LD) clumping. The approach iterated from the first to last SNP on each chromosome sorted from smallest (most significant) to largest p-value (less significant) in the discovery sample that had not already been clumped (denoting this as the index SNP) and formed clumps of all other SNPs that were within 10 Mb of and not in LD with the index SNP ($r^2 > 0.1$) based on HapMap2 CEU genotype data. In total, for autosomal SNPs including the HLA region, there were 24,722 SNPs clumped using the EGMA as a discovery sample, and 24,642 SNPs clumped using the MGMA as a discovery sample. For autosomal SNPs excluding the HLA region, there were 24,688 SNPs clumped using the EGMA as a discovery sample, and 24,621 SNPs clumped using the MGMA as a discovery sample. To test for concordant SNP effects between the two datasets, a total of 144 key SNP subsets were examined, each generated using twelve P-value thresholds ($P = 0.01$, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 and 1.0). Fisher's exact test was applied to test whether the overlapping effects were in the same direction after conditioning on their GWAS p-value, where a Fisher's exact test p-value ($P_{FT} < 0.05$) means the SNP effects are nominally correlated. 1,000 permutations were performed to examine the significance of observing a specific

proportion of subsets with correlated SNP effects—adjusted for testing all 144 subsets ($P_{FTsig\text{-}permuted}$)[27].

2.3.2.5 Genetic risk score analysis (GRS)

To further validate the SECA results, genetic risk scores (GRS)[28] were constructed using the independent shared SNPs separated by <10 Mb that are not in linkage disequilibrium ($r^2 < 0.1$) and selected as having the smallest p-value in the discovery sample with a nominally significant p-value <0.05, and in the target sample with p-value <1.0 (the different genotyping platforms between the two datasets had no impact on the analysis). GRS analysis requires individual-level GWAS SNP data for the target sample. As only part of the genotype data for the EGMA and MGMA could be obtained, the first GRS analysis (EGMA as a discovery sample) used 5,862 clumped SNPs to generate EGMA-based genetic risk scores in the MS GWAS obtained from dbGaP (phs000275, phs000171 and phs000139) and ANZgene (cases = 3,252; controls = 5,725).For details on data cleaning, please refer to our previous paper[29]. The second GRS analysis (MGMA as a discovery sample) used 5,770 clumped SNPs to generate MGMA-based genetic risk scores in the QIMR Twin Families EBNA-1 GWAS. For binary disease traits (cases and controls), the GRS was used as a predictor to test whether higher mean GRS were observed for cases (MS) than for controls in the target samples. For quantitative traits, the GRS was used to test the correlation between GRS and the target phenotype (anti-EBNA-1 IgG titers).

2.3.2.6 Joint EBNA-1 and MS Meta-Analysis

To identify genetic risk factors contributing to both anti-EBNA-1 IgG titer and MS risk outside the HLA region, a joint meta-analysis of EGMA results and MGMA results was carried out using a sample size-weighted method based on p-values in METAL[30]. The power analysis showed that to achieve 80% power for detecting a common variant with OR of a modest magnitude (OR=1.15) would require N=10,000 samples.

2.3.2.7 Gene-gene interaction analysis

To determine whether interactions existed between the loci that were found to contribute to both anti-EBNA-1 IgG titer and MS risk, we defined two logistic regression models; one took into account only the marginal effects of both SNPs and the other was a full logistic regression model including a pairwise interaction between the two SNPs. A likelihood ratio test with four degrees of freedom was then used to determine whether there was any statistical difference between the two models [31]. The datasets used for the analysis were the MS GWAS obtained from dbGaP and ANZgene listed above (cases = 3,252; controls = 5,725).

2.4 Results

2.4.1 HLA region as major quantitative trait loci influencing anti-EBNA-1 IgG titer

The distributions of anti-EBNA-1 IgG titers using four groups (<250; 250-1,000; 1,001-2,000; >2,000) in different age ranges are shown in Figure 2.2. As age

increased, the percentage of individuals with higher anti-EBNA-1 IgG titers increased dramatically. Heritability for anti-EBNA-1 IgG titers was 42%. The HLA region on chromosome 6 contained multiple SNPs exceeding the threshold for a significant association with anti-EBNA-1 IgG titers ($P < 5\times10^{-8}$) (Supplementary Figure 2.2). A Q-Q plot is shown in Supplementary Figure 2.3. Of these HLA SNPs, the most significant imputed SNP was rs9268923 ($P = 1.22\times10^{-11}$), located between the genes *HLA-DRA* and *HLA-DRB5*. The most significant genotyped SNP was rs2516049 ($P = 4.11\times10^{-9}$; Table 2.1), located between the genes *HLA-DRB1* and *HLA-DQA1*.

Figure 2.2: Anti-EBNA-1 IgG titer distribution in different age ranges for twins and their parents. Anti-EBNA-1 IgG titers are divided into four groups (<250, 250-1000,1001-2000, >2000).

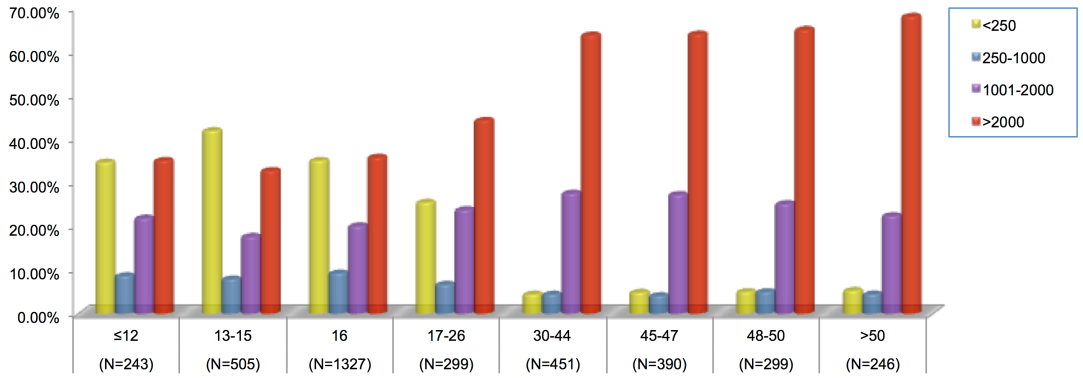


Figure 2.3: Association analysis of HLA region and EBNA-1 titers in QIMR Twin Families EBNA-1 GWAS (QTFEGWAS)(N=3,599). Results are shown (a) unconditional analysis. (b) conditional analysis on rs9268923. (c) conditional analysis on rs9268923 and rs2267647. The top significant genotyped SNP rs2516049 was also highlighted in red. SNPs with a P value $<1.0x10^{-5}$ are highlighted in green.

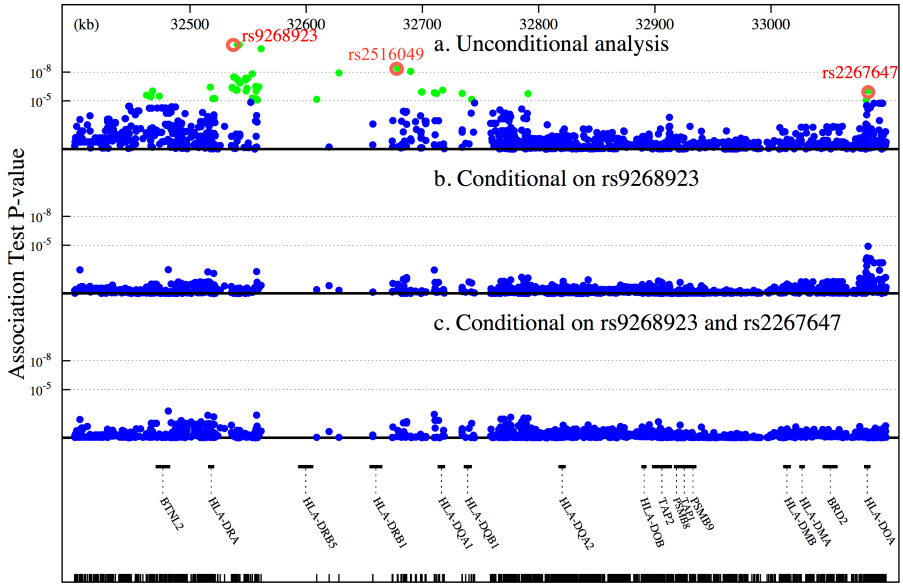


Table 2.1 : Genome-wide association analysis results of QIMR twin families. Here E(A) refers to the effect allele. O(A) refers to the opposite allele. Score refers to the imputed or genotype score of the SNP. Freq refers to the E(A) frequency. Effect corresponds to standard deviation units for the transformed phenotype. Children[1] refers to using all the children samples, while children[2] refers to only for anti-EBNA-1 IgG titers positive children.

							Parents+children(N=3,599)			Parents(N=1,243)		Children[1](N=2,356)		Chidren[2](N=1,488)	
SNP	CHR	POS	E(A)	O(A)	SNP Type	Score	**Freq**	**Effect**	**P**	**Effect**	**P**	**Effect**	**P**	**Effect**	**P**
rs9268923	6	32540813	C	T	Imputed	0.905	0.552	0.180	1.22×10^{-11}	0.252	1.37×10^{-13}	0.148	2.93×10^{-05}	0.225	9.72×10^{-12}
rs9268969	6	32542327	C	T	Imputed	0.905	0.552	0.180	1.22×10^{-11}	0.252	1.37×10^{-13}	0.148	2.93×10^{-05}	0.225	9.72×10^{-12}
rs6916742	6	32561169	C	T	Imputed	0.879	0.569	-0.176	3.12×10^{-11}	-0.200	2.40×10^{-09}	-0.160	5.95×10^{-06}	-0.215	5.05×10^{-11}
rs2516049	6	32678378	T	C	Genotyped	0.999	0.666	0.162	4.11×10^{-09}	0.210	1.31×10^{-09}	0.142	1.36×10^{-04}	0.180	1.68×10^{-07}
rs522308	6	32689900	C	T	Imputed	0.989	0.664	0.159	8.56×10^{-09}	0.207	1.89×10^{-09}	0.138	2.03×10^{-04}	0.178	2.36×10^{-07}
rs11757159	6	32628250	C	T	Imputed	0.972	0.636	0.154	1.24×10^{-08}	0.206	1.44×10^{-09}	0.130	3.59×10^{-04}	0.170	4.07×10^{-07}
rs9405112	6	32553578	G	A	Imputed	0.977	0.662	0.156	1.57×10^{-08}	0.215	8.08×10^{-10}	0.128	5.30×10^{-04}	0.176	2.38×10^{-07}
rs9268853	6	32537621	T	C	Imputed	0.989	0.657	0.152	3.30×10^{-08}	0.206	3.36×10^{-09}	0.125	6.96×10^{-04}	0.172	4.64×10^{-07}
rs9268858	6	32537736	T	C	Imputed	0.989	0.657	0.152	3.30×10^{-08}	0.206	3.36×10^{-09}	0.125	6.96×10^{-04}	0.172	4.64×10^{-07}
rs9268877	6	32539125	G	A	Imputed	0.977	0.585	-0.146	3.35×10^{-08}	-0.184	3.46×10^{-08}	-0.130	2.33×10^{-04}	-0.176	7.00×10^{-08}
rs7739203	6	32549533	G	A	Imputed	0.973	0.585	-0.145	3.90×10^{-08}	-0.187	1.99×10^{-08}	-0.129	2.65×10^{-04}	-0.175	8.13×10^{-08}
rs7739357	6	32549619	G	A	Imputed	0.973	0.585	-0.145	3.90×10^{-08}	-0.187	1.99×10^{-08}	-0.129	2.65×10^{-04}	-0.175	8.13×10^{-08}

As the HLA region is genetically complex, with a low rate of recombination and long-range linkage disequilibrium, we carried out two conditional analyses to check whether there were independent peaks, first by conditioning on rs9268923, and then on both rs9268923 and rs2267647 (Figure 2.3). However, after conditional analysis, no other SNPs showed a significant association, suggesting these top SNPs are within one haplotype block that contains all the variants modulating anti-EBNA-1 IgG titers.

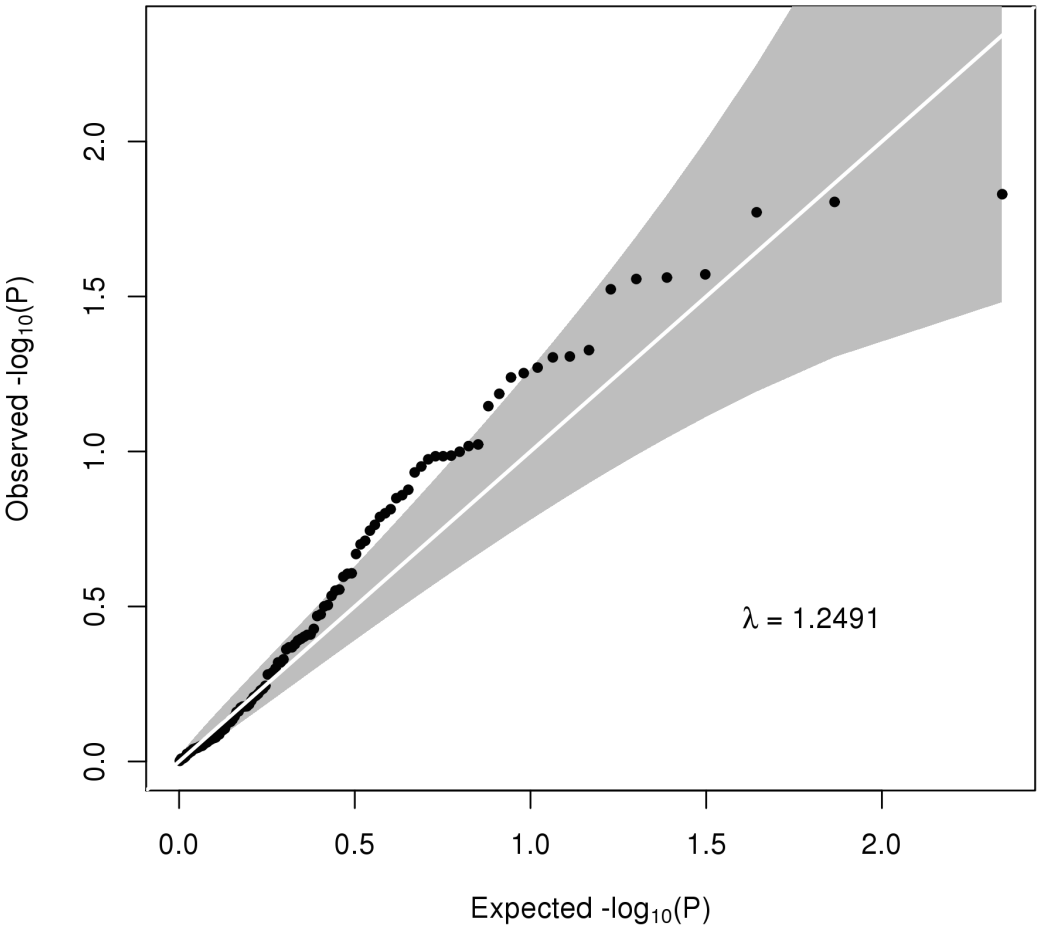
Association analyses were also conducted separately for parents and children (Supplementary Table 2.1:). Analysis of parents only (N = 1,243), implicated the same SNPs in the HLA region (Supplementary Figure 2.2), that were found using all samples: imputed SNP rs9268923 ($P = 1.37\times10^{-13}$) and genotyped SNP rs2516049 ($P = 1.31\times10^{-9}$; Table 2.1). In contrast, analysis of all children (N = 2,356), did not identify any significant SNPs in the HLA region (Supplementary Figure 2.2). However, once we restricted analyses to only EBNA-1 positive (>300 units/ml) children (N = 1,488), we again observed the same significant associations in the HLA region (Supplementary Figure 2.2). Q-Q plots are shown in Supplementary Figure 2.3 respectively.

In the EBNA-1 GWA meta-analysis (EGMA) (inflation factor, $\lambda = 1.017$), 45 SNPs that reached genome-wide significance were located in the HLA region ($P < 5.0\times10^{-8}$; Supplementary Table 2.2). Of these, rs2516407 remained the most significant ($P = 3.32\times10^{-20}$). No other loci were genome-wide significant outside the HLA region (Supplementary Figure 2.4). The inflation factors (λ) for the analyses were close to 1, suggesting population stratification had minimal influence on test statistical distribution; therefore, we did not adjust for genomic control in our analysis.

2.4.2 Anti-EBNA-1 IgG titers as possible risk factors for the development of MS: the effect of HLA and non-HLA region genes

The Q-Q plot (Figure 2.4) shows a significant excess of smaller EGMA p- values for the 110 non-HLA MS SNPs recently published by the International Multiple Sclerosis Genetics Consortium [24] than expected by chance ($\lambda = 1.25$). In SNP effect concordant analysis, using the EGMA as a discovery sample, the clumped SNP effects were positively correlated with MGMA effects, either including ($P_{FTsig\text{-}permuted} = 0.026$; 95% CI: 0.018 – 0.038) or excluding ($P_{FTsig\text{-}permuted} = 0.049$; 95% CI: 0.037 – 0.064) the HLA region (Supplementary Figure 2.5). However, using the MGMA as a discovery sample, the clumped SNP effects were not positively correlated with the EGMA SNP effects, either including ($P_{FTsig\text{-}permuted} = 0.173$; 95% CI: 0.151 – 0.197) or excluding ($P_{FTsig\text{-}permuted} = 0.157$; 95% CI: 0.136 – 0.181) the HLA region (Supplementary Figure 2.6).

Figure 2.4: QQ plot for 110 non-HLA MS SNPs in the EBNA-1 GWAS Meta-Analysis (EGMA). The 95% confidence envelope (shaded grey) was formed by calculating, for each order statistic (p-values ranked from smallest to largest), the 2.5th and 97.5th centiles of the distribution of the order statistic under random sampling and the null hypothesis.

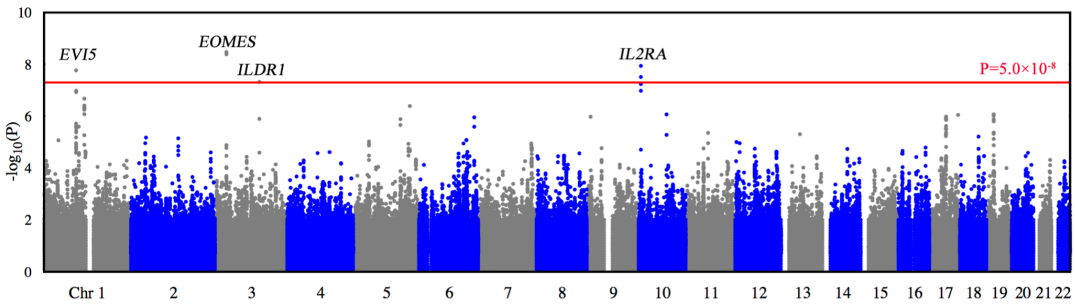


The SNP effect concordant analysis results (excluding the HLA) were further validated by genetic risk score analysis. A significantly higher GRS, constructed using the EGMA as a discovery sample, was observed in MS cases than controls in

the target dbGaP and ANZgene GWAS (P = 0.007; Supplementary Figure 2.7). In contrast, a GRS constructed using MGMA as a discovery sample, did not correlate with anti-EBNA-1 IgG titers in the target QIMR Twin Families EBNA-1 GWAS samples (P = 0.270, correlation coefficient = -0.030; Supplementary Figure 2.8). Altogether, these results indicate that the genetic risk for elevated anti-EBNA-1 IgG titer is positively correlated with increased MS susceptibility, whereas the reverse does not hold. That is, the genetic risk for MS is not positively correlated with elevated anti-EBNA-1 IgG titer level.

The combined meta-analysis of the EGMA and MGMA results (excluding the HLA region) (λ = 1.046) showed several SNP associations that reached genome-wide significance ($P < 5.0\times10^{-8}$; Supplementary Table 2.3). The approximate chromosomal location of the significant SNPs are shown in the Manhattan plot in Figure 2.5 and labelled with the nearest gene.

Figure 2.5 Manhattan plot for the pooled meta-analysis of EBNA-1 GWAS Meta-Analysis (EGMA) and MS GWAS Meta-Analysis (MGMA) (excluding the HLA region). The 22 autosomal chromosomes separated by two different colors are displayed on the X-axis. The -log$_{10}$(P) values are displayed on the Y-axis. Genome-wide association significance level is shown as the red line (P<5.0x10^{-8}).



These genes include *EVI5*(1p22.1)*, EOMES*(3p24.1)*, ILDR1*(3q13.33) and *IL2RA*(10p15.1). For detailed regional plots, see Supplementary Figure 2.9. Of these, the most interesting gene is *Ecotropic viral integration site 5* (*EVI5*) on chromosome 1, in which a missense mutation SNP rs11808092 ($P = 1.71\times10^{-8}$) results in an amino acid change from glutamine to histidine. The impact of this amino acid substitution on the structure and function of the protein product was predicted as benign using PolyPhen-2[32], with a score of 0.003 (sensitivity: 0.98; specificity: 0.44). SIFT[33] gave similar results (tolerated), whilst MutationAssessor[34] scored it as "medium" (2.255).

In the interaction analysis, tagged SNPs within each genome-wide significant peak were selected ($r^2 > 0.65$). In total we studied five SNPs: rs11808092 (1p22.1), rs427221 (3p24.1), rs2255214 (3q13.33), rs2516049 (6p21.32), and rs12722561 (10p15.1). We observed a suggestive synergistic interaction between rs2516049 and rs11808092 ($P = 0.006$; Supplementary Table 2.4, Supplementary Table 2.5).

2.5 Discussion

We have undertaken a large GWAS of anti-EBNA-1 IgG titers and provide strong evidence that the HLA region serves as the main quantitative trait locus for anti-EBNA-1 IgG titers. Our results strongly suggest that there are shared genetic risk factors that influence both anti-EBNA-1 IgG titers and MS risk. Our results further indicate that the genetic risk for elevated anti-EBNA-1 IgG titer is positively correlated with increased MS susceptibility but the reverse importantly is not supported, i.e, MS risk genetic differences do not increase anti-EBNA-1 titers. In our study the heritability of anti EBNA1 titers is 42% clearly suggesting that other factors besides genetic differences are also important in MS pathogenesis and that there is

still a clear strong association with prior EBV infection in MS risk modulated by the individual's genetic make-up.

In our results the strongest non-HLA EBNA-1 association was found for *EVI5*. The major function of *EVI5* is the modulation of cell cycle progression, cytokinesis and cellular membrane traffic[35]. Several studies have described the involvement of *EVI5* in the risk and severity of MS using case-control and cross-sectional study designs. Using the candidate gene approach, Hoppenbrouwers et al found that SNPs within *EVI5* showed a moderate association with MS risk (OR = 1.90 – 2.01; $P = 0.01$)[36]. The ANZgene consortium, found a significant interaction between SNPs in *EVI5* and *HLA-DR15* in predicting MS risk ($P = 0.001$)[37]. This finding is supported by recent work demonstrating that *EVI5* genotype is associated with a greater odds of having a more severe relapse among individuals who carry the risk allele of *HLA-DRB1*1501*[38]. These studies showed associations with MS parameters, but did not consider genetic effects in concert with anti-EBNA-1 IgG levels. Direct evidence of a link between EBV infection and gene expression changes has been provided by the work of van Aalderen et al.[39], who found that *EOMES* expression was altered after EBV infection. While in another study by Parnell et al.[40] analyzing the expression of MS associated transcription factors using whole blood, *EOMES* expression was significantly lower in MS cases compared with healthy controls.

In addition to the correlation in genetic risk between anti-EBNA-1 IgG titer and MS, the observed excess of smaller EBNA-1 GWA meta-analysis p-values for the recently reported 110 non-HLA MS risk SNPs[24], suggests that other genetic factors likely influence the interaction between MS risk and EBV infection. Hence, analysis of larger EBNA-1 GWAS datasets and/or analysis of other markers of EBNA immune

responses (e.g., EBNA-2 or EBNA-3 titers) using the healthy population should provide a novel opportunity to identify new and characterize existing genetic risk factors for MS. Our findings indicate that these studies offer a powerful approach to identify the specific immune responses components that contribute most to MS susceptibility. Indeed, considering the high cost and difficulty in amassing larger samples of MS cases, we suggest focus could switch to performing large studies; such as these, aimed at identifying genetic loci influencing EBNA immune response. Such studies should also provide insight into other EBV-associated autoimmune diseases or cancers such as Burkitt's lymphoma, nasopharyngeal carcinoma.

In conclusion, our results suggest that genetic factors influencing differential immune system reactivity to EBV infection increase the risk of MS in genetically predisposed individuals exposed to EBV, most notably in adolescence or adult life.

2.6 Summary

Background: Infection with the Epstein-Barr virus (EBV) is associated with an increased risk of multiple sclerosis (MS).

Objective: We sought genetic loci influencing EBV nuclear antigen-1 (EBNA-1) IgG titers and hypothesized that they may play a role in MS risk.

Methods: We performed a genome-wide association study (GWAS) of anti-EBNA-1 IgG titers in 3,599 individuals from an unselected twin family cohort, followed by a meta-analysis with data from an independent EBNA-1 GWAS. We then examined the shared polygenic risk between the EBNA-1 GWAS (effective sample size (N_{eff}) = 5,555) and a large MS GWAS (N_{eff} = 15,231).

Results: We identified one locus of strong association within the human leukocyte antigen (HLA) region, of which the most significantly associated genotyped SNP was rs2516049 ($P=4.11x10^{-9}$). A meta-analysis including data from another EBNA-1 GWAS in a cohort of Mexican-American families confirmed that rs2516049 remained the most significantly associated SNP ($P=3.32x10^{-20}$). By examining the shared polygenic risk, we show that the genetic risk for elevated anti-EBNA-1 titers is positively correlated with the development of MS, and that elevated EBNA-1 titers are not an epiphenomena secondary to MS. In the joint meta-analysis of EBNA-1 titers and MS, loci at 1p22.1, 3p24.1, 3q13.33 and 10p15.1 reached genome-wide significance ($P < 5\times10^{-8}$).

Conclusions: Our results suggest that apart from the confirmed HLA region, the association of anti-EBNA-1 IgG titer with MS risk is also mediated through non-HLA genes, and that studies aimed at identifying genetic loci influencing EBNA immune response provide a novel opportunity to identify new and characterize existing genetic risk factors for MS.

2.7 References

1. Zhou Y, Simpson S, Jr., Holloway AF, Charlesworth J, van der Mei I, Taylor BV. The potential role of epigenetic modifications in the heritability of multiple sclerosis. *Mult. Scler.* 2014;20(2):135-140.

2. Lucas RM, Hughes AM, Lay ML, et al. Epstein-Barr virus and multiple sclerosis. *J. Neurol. Neurosurg. Psychiatry.* 2011;82(10):1142-1148.

3. Young LS, Rickinson AB. Epstein-Barr virus: 40 years on. *Nat. Rev. Cancer.* 2004;4(10):757-768.

4. Pender MP. CD8+ T-Cell Deficiency, Epstein-Barr Virus Infection, Vitamin D Deficiency, and Steps to Autoimmunity: A Unifying Hypothesis. *Autoimmune Dis.* 2012;2012:189096.

5. Simpson Jr. S, Blizzard L, Otahal P, van der Mei I, Taylor B. Latitude is significantly associated with the prevalence of multiple sclerosis: a meta-analysis. *J. Neurol. Neurosurg. Psychiatry.* 2011;82(10):1132-1141.

6. Pakpoor J, Disanto G, Gerber JE, et al. The risk of developing multiple sclerosis in individuals seronegative for Epstein-Barr virus: a meta-analysis. *Mult. Scler.* 2013;19(2):162-166.

7. Munger KL, Levin LI, O'Reilly EJ, Falk KI, Ascherio A. Anti-Epstein-Barr virus antibodies as serological markers of multiple sclerosis: a prospective study among United States military personnel. *Mult. Scler.* 2011;17(10):1185-1193.

8. Strautins K, Tschochner M, James I, et al. Combining HLA-DR risk alleles and anti-Epstein-Barr virus antibody profiles to stratify multiple sclerosis risk. *Mult. Scler.* 2014;20(3):286-294.

9. Lunemann JD, Jelcic I, Roberts S, et al. EBNA1-specific T cells from patients with multiple sclerosis cross react with myelin antigens and co-produce IFN-gamma and IL-2. *J. Exp. Med.* 2008;205(8):1763-1773.

10. Cepok S, Zhou D, Srivastava R, et al. Identification of Epstein-Barr virus proteins as putative targets of the immune response in multiple sclerosis. *J. Clin. Invest.* 2005;115(5):1352-1360.

11. Lunemann JD, Tintore M, Messmer B, et al. Elevated Epstein-Barr virus-encoded nuclear antigen-1 immune responses predict conversion to multiple sclerosis. *Ann. Neurol.* 2010;67(2):159-169.

12. Farrell RA, Antony D, Wall GR, et al. Humoral immune response to EBV in multiple sclerosis is associated with disease activity on MRI. *Neurology.* 2009;73(1):32-38.

13. Levin LI, Munger KL, Rubertone MV, et al. Temporal relationship between elevation of Epstein-Barr virus antibody titers and initial onset of neurological symptoms in multiple sclerosis. *JAMA : the journal of the American Medical Association.* 2005;293(20):2496-2500.

14. Lucas RM, Ponsonby AL, Dear K, et al. Current and past Epstein-Barr virus infection in risk of initial CNS demyelination. *Neurology.* 2011;77(4):371-379.

15. Levin LI, Munger KL, O'Reilly EJ, Falk KI, Ascherio A. Primary infection with the Epstein-Barr virus and risk of multiple sclerosis. *Ann. Neurol.* 2010;67(6):824-830.

16. Ascherio A, Munger KL, Lunemann JD. The initiation and prevention of multiple sclerosis. *Nat. Rev. Neurol.* 2012;8(11):602-612.

17. Rubicz R, Yolken R, Drigalenko E, et al. A genome-wide integrative genomic study localizes genetic factors influencing antibodies against Epstein-Barr virus nuclear antigen 1 (EBNA-1). *PLoS Genet.* 2013;9(1):e1003147.

18. Zhu G, Montgomery GW, James MR, et al. A genome-wide scan for naevus count: linkage to CDKN2A and to other chromosome regions. *Eur. J. Hum. Genet.* 2007;15(1):94-102.

19. Wright. MJ, Martin NG. Brisbane Adolescent Twin Study: Outline of study methods and research projects. *Aust. J. Psychol.* 2004;56(2):65-78.

20. Ferreira MA, Mangino M, Brumme CJ, et al. Quantitative trait loci for CD4:CD8 lymphocyte ratio are associated with risk of type 1 diabetes and HIV-1 immune control. *Am. J. Hum. Genet.* 2010;86(1):88-92.

21. Abecasis GR, Cherny SS, Cookson WO, Cardon LR. Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. *Nat. Genet.* 2002;30(1):97-101.

22. Almasy L, Blangero J. Multipoint quantitative-trait linkage analysis in general pedigrees. *Am. J. Hum. Genet.* 1998;62(5):1198-1211.

23. Liu JZ, Tozzi F, Waterworth DM, et al. Meta-analysis and imputation refines the association of 15q25 with smoking quantity. *Nat. Genet.* 2010;42(5):436-440.

24. International Multiple Sclerosis Genetics Consortium, Beecham AH, Patsopoulos NA, et al. Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. *Nat. Genet.* 2013;45(11):1353-1360.

25. Lee D, Bigdeli TB, Riley BP, Fanous AH, Bacanu SA. DIST: direct imputation of summary statistics for unmeasured SNPs. *Bioinformatics.* 2013;29(22):2925-2927.

26. Patsopoulos NA, Bayer Pharma MSGWG, Steering Committees of Studies Evaluating I-b, et al. Genome-wide meta-analysis identifies novel multiple sclerosis susceptibility loci. *Ann. Neurol.* 2011;70(6):897-912.

27. Nyholt DR. SECA: SNP effect concordance analysis using genome-wide association summary results. *Bioinformatics.* 2014;30(14):2086-2088.

28. International Schizophrenia Consortium, Purcell SM, Wray NR, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature.* 2009;460(7256):748-752.

29. Lin R, Charlesworth J, Stankovich J, et al. Identity-by-descent mapping to detect rare variants conferring susceptibility to multiple sclerosis. *PLoS One.* 2013;8(3):e56379.

30. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics.* 2010;26(17):2190-2191.

31. Emily M, Mailund T, Hein J, Schauser L, Schierup MH. Using biological networks to search for interacting loci in genome-wide association studies. *Eur. J. Hum. Genet.* 2009;17(10):1231-1240.

32. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nat Methods.* 2010;7(4):248-249.

33. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.* 2009;4(7):1073-1081.

34. Reva B, Antipin Y, Sander C. Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Res.* 2011;39(17):e118.

35. Lim YS, Tang BL. The Evi5 family in cellular physiology and pathology. *FEBS Lett.* 2013;587(12):1703-1710.

36. Hoppenbrouwers IA, Aulchenko YS, Ebers GC, et al. EVI5 is a risk gene for multiple sclerosis. *Genes Immun.* 2008;9(4):334-337.

37. Australia New Zealand Multiple Sclerosis Genetics Consortium. Genome-wide association study identifies new multiple sclerosis susceptibility loci on chromosomes 12 and 20. *Nat. Genet.* 2009;41(7):824-828.

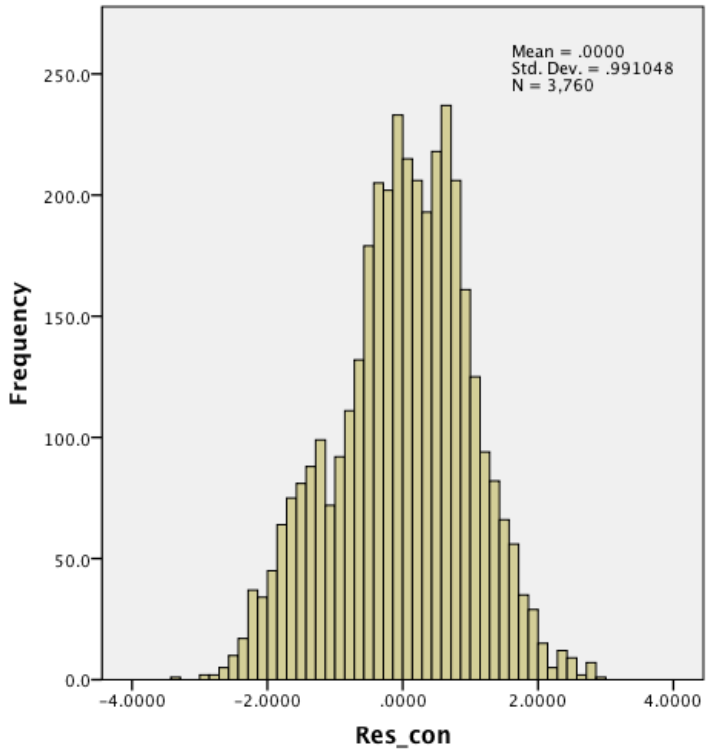
38. Mowry EM, Carey RF, Blasco MR, et al. Multiple sclerosis susceptibility genes: associations with relapse severity and recovery. *PLoS One.* 2013;8(10):e75416.

39. van Aalderen MC, Remmerswaal EB, Verstegen NJ, et al. Infection history determines the differentiation state of human CD8+ T cells. *J. Virol.* 2015;89(9):5110-5123.

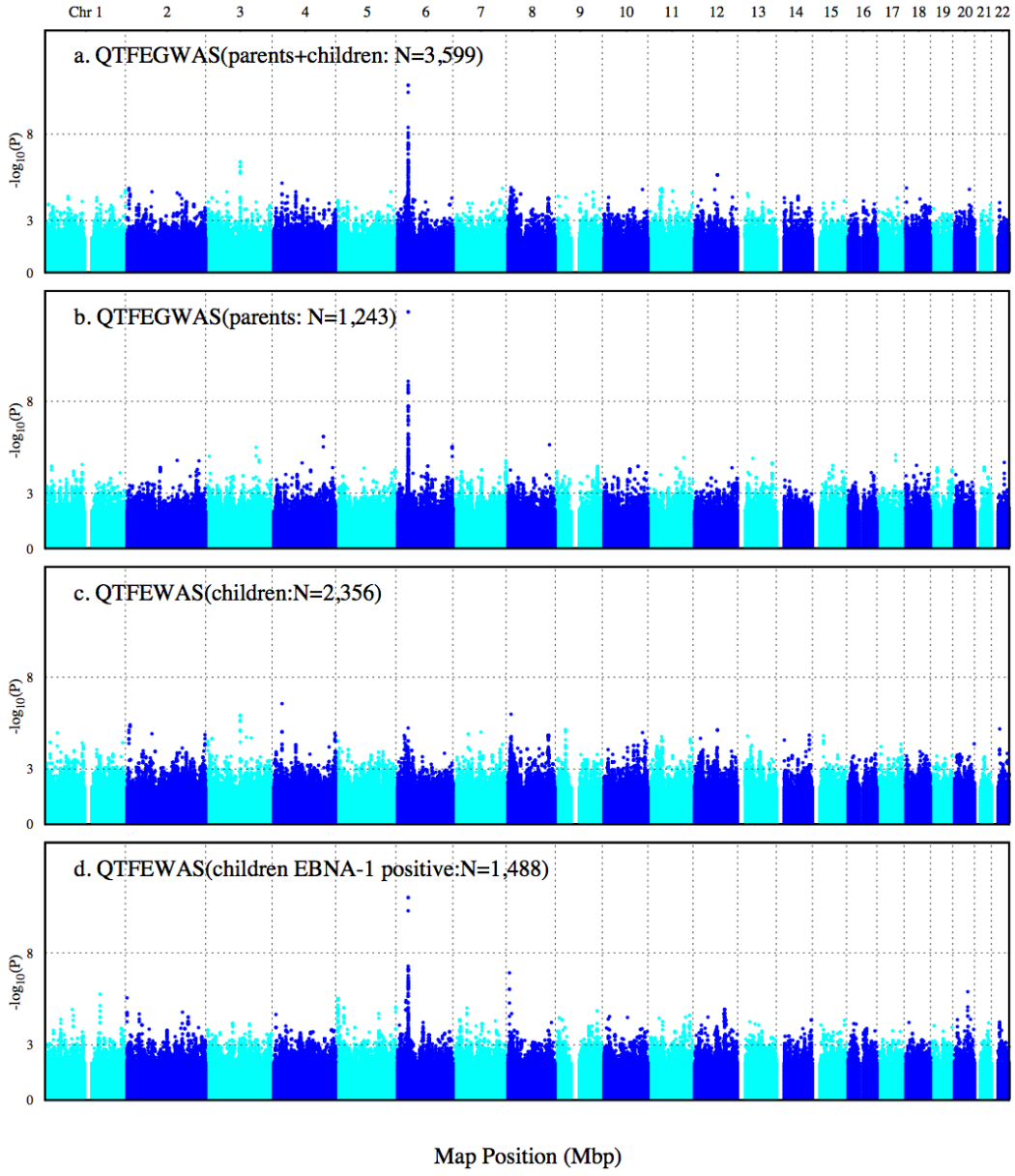
40. Parnell GP, Gatt PN, Krupa M, et al. The autoimmune disease-associated transcription factors EOMES and TBX21 are dysregulated in multiple

sclerosis and define a molecular subtype of disease. *Clin. Immunol.* 2014;151(1):16-24.

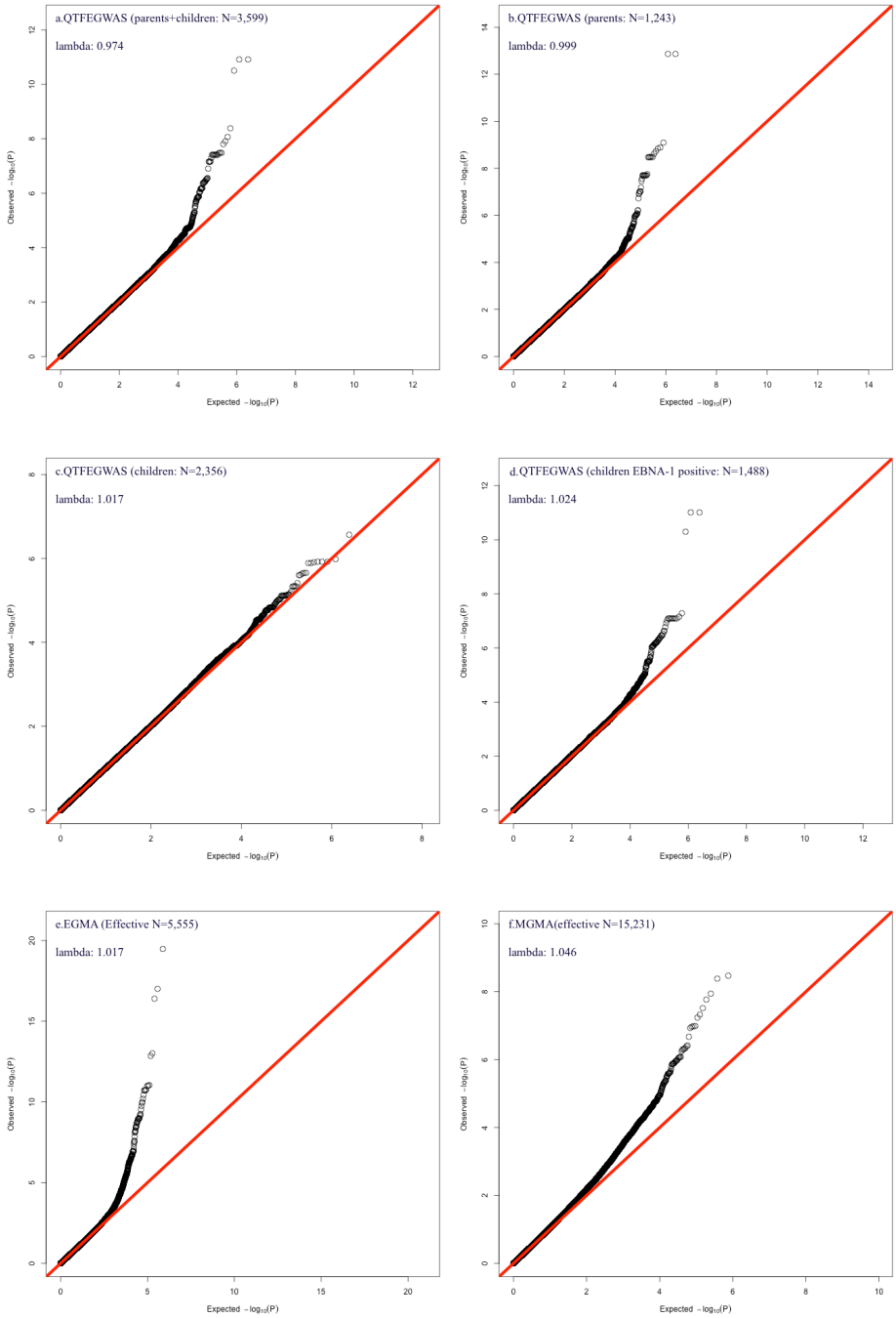
Supplementary Figure 2.1 : Distribution of standardized residuals for continuous anti-EBNA-1 titers in the QIMR sample after generalized linear model (GLM) analysis with box effect, row effect, and column effect as fixed factors while also controlling sex, age, age x sex, age^2 and age^2 x sex. Here the Res_con refer to the residual value of anti-EBNA-1 titres, while Frequency refer to the number of samples.



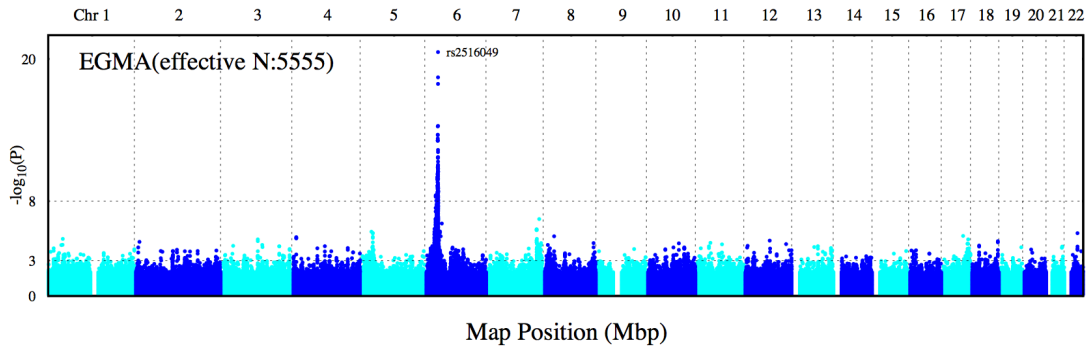
Supplementary Figure 2.2: Manhattan Plot for QIMR Twin Families EBNA-1 GWAS (QTFEGWAS) a) using all samples (N = 3,599). b) only parents (N = 1,243). c) only children (N = 2,356).d) only children with positive anti-EBNA-1 IgG titres (N = 1,488)



Supplementary Figure 2.3: QQ plot for QTFEGWAS using a) both parents and children (N = 3,599). b) only parents (N = 1,243). c) only children (N = 2,356). d) only children with positive anti-EBNA-1 IgG titres (N = 1,488) e)QQ plot for EGMA (effective N = 5,555). f) QQ plot for MGMA excluding HLA region (chr6: 25Mb-35Mb) (effective N = 20,786).

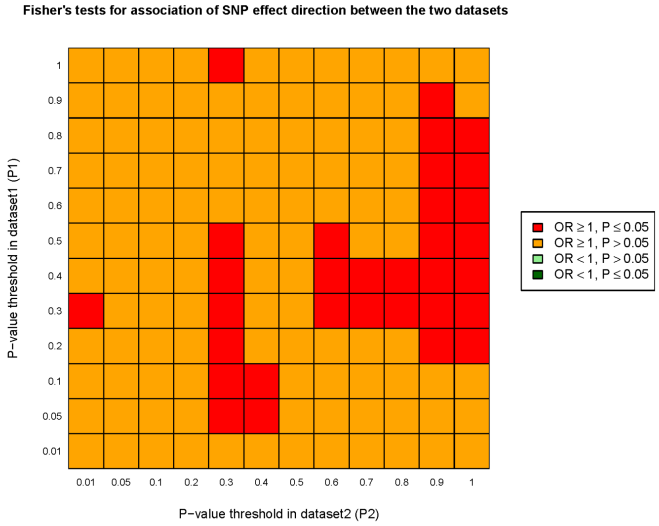


Supplementary Figure 2.4 : Manhattan plot for Meta-Analysis of EBNA-1 IgG GWAS fixed effect (EGMA): QIMR Twin Families EBNA-1 GWAS (QTFEGWAS) (N = 3,599) and Mexican-American families EBNA-1 GWAS (MAFEGWAS) (N = 1,956).

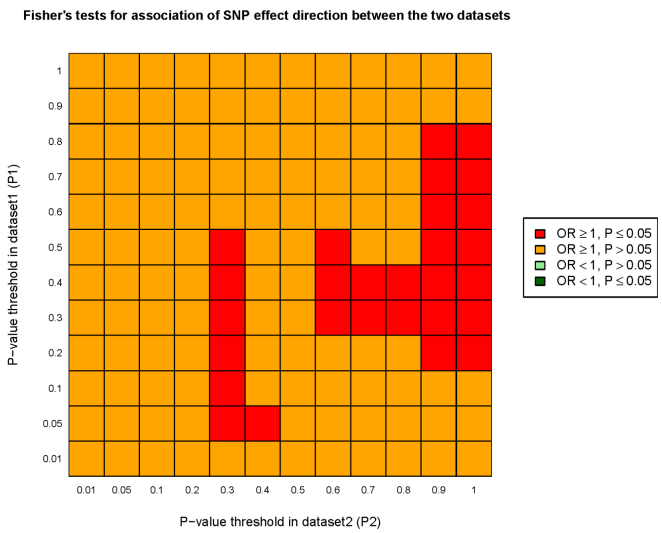


Supplementary Figure 2.5: Heatmap for SECA results including/ excluding HLA region using EBNA-1 GWAS Meta analysis (EGMA) as the discovery sample (dataset1), and the MS GWAS Meta analysis (MGMA) as the target sample (dataset2). $P_{FTsig\text{-}permuted}$ refers to permutation p value for the observed significant Fisher's test . The four colors represented all the 4 outcomes for the Fisher's tests(OR>=1,P<=0.05; OR>=1,P>=0.05; OR<=1,P<=0.05; OR<=1,P>=0.05)

a. Including HLA region. $P_{FTsig\text{-}permuted}$= 0.026 (95% CI: 0.018-0.038)

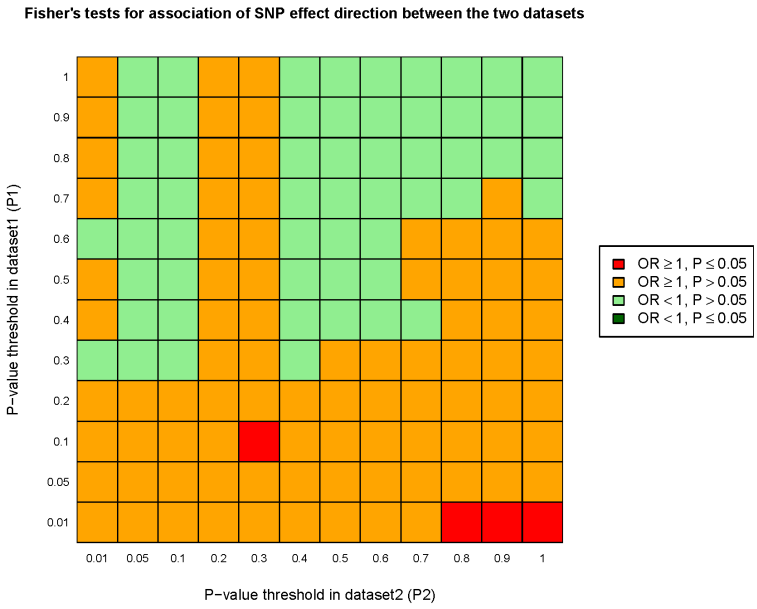


b. Excluding HLA region. $P_{FTsig\text{-}permuted}$=0.049 (95% CI: 0.037-0.064)

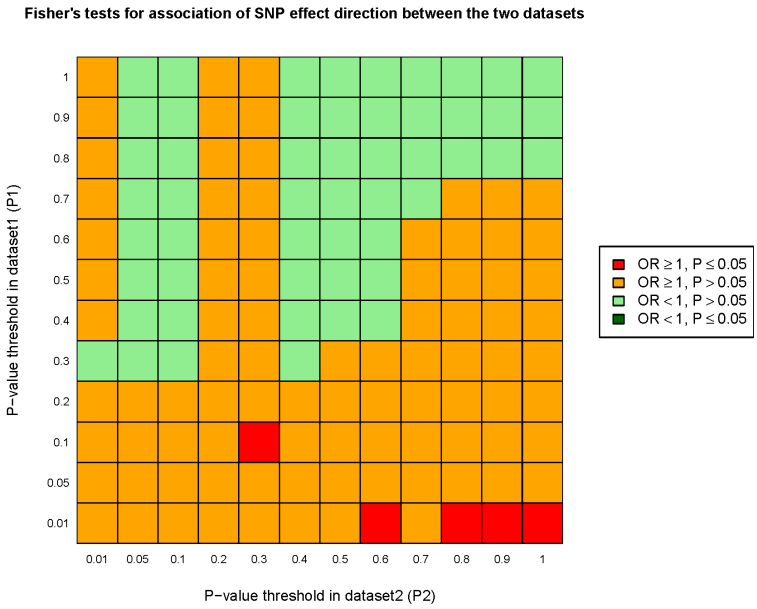


Supplementary Figure 2.6: Heatmap for SECA results including/ excluding HLA region using the MS GWAS Meta analysis (MGMA) as the discovery sample (dataset1), and the EBNA-1 GWAS Meta analysis (EGMA) as the target sample (dataset2).

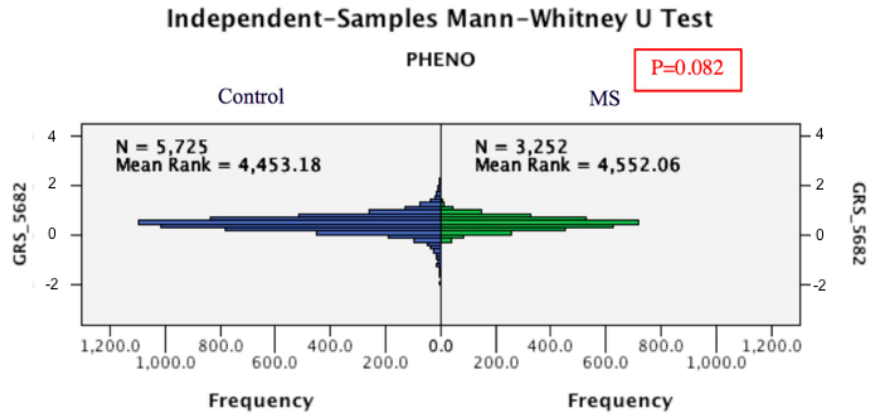
a. Including HLA region. $P_{FTsig\text{-}permuted}$ = 0.173 (95% CI: 0.151-0.197)

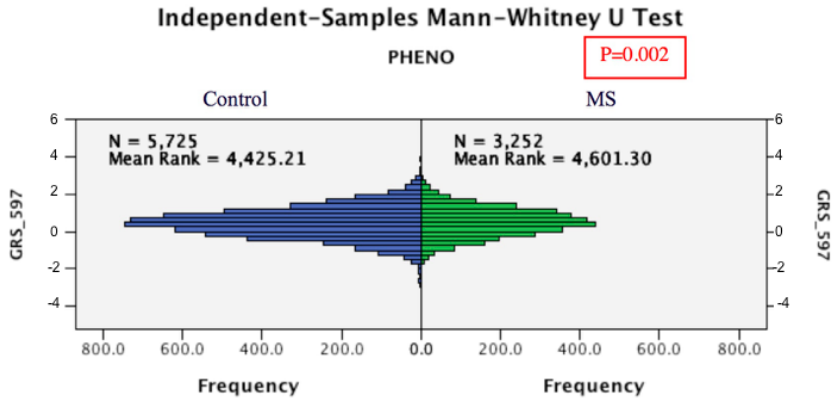


b. Excluding the HLA region. $P_{FTsig\text{-}permuted}$= 0.157 (95% CI: 0.136-0.181)

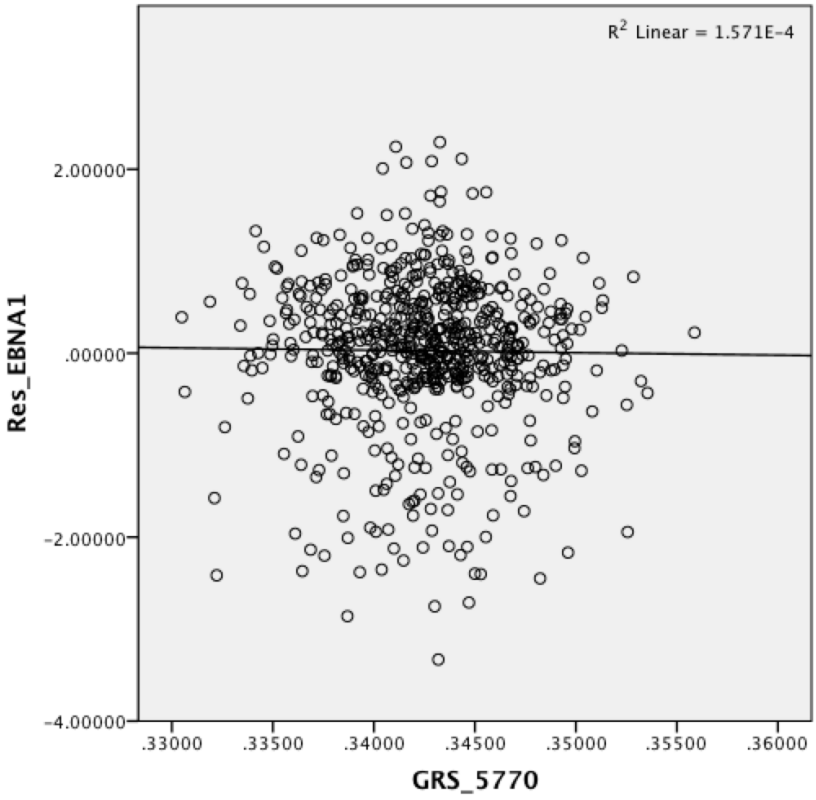


Supplementary Figure 2.7: Genetic Risk Score (GRS) results using the EBNA-1 GWAS meta analysis (EGMA) as the discovery sample, and dbGaP and ANZgene (MS cases) as target samples. a) The EBNA-1 constructed GRS values (5,682 SNPs with $P_{EGMA} \leq 0.05$) are higher in MS than controls and this difference is significant using an independent sample t-test (p=0.007) and borderline significant using a nonparametric test (p=0.082). b) Using the enriched SNP set (597 SNPs with $P_{EGMA} \leq 0.05$ and $P_{MGMA} \leq 0.1$) for which the correlation of SNP effects is strongest, the result is more significant whether using an independent sample t test (p=0.003) or nonparametric test (p=0.002). Below are the plot results for the nonparametric test. The mean rank value was used to compare which was higher. A.



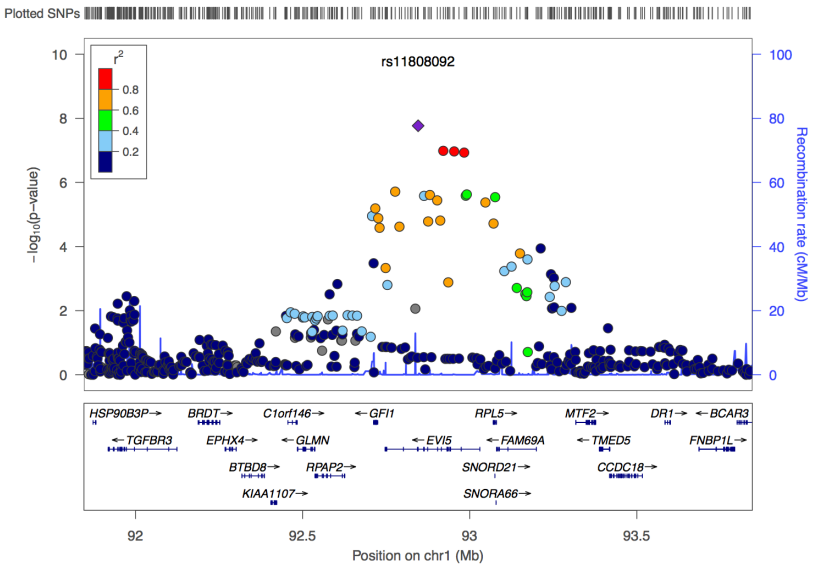


Supplementary Figure 2.8: Genetic risk score (GRS) results using the MS GWAS meta-analysis (MGMA) as the discovery sample, and the QIMR Twin Families EBNA-1 GWAS as a target sample. The MS constructed GRS values (5,770 SNPs with $P_{MGMA} \leq 0.05$) do not correlate with the anti-EBNA-1 titres in QIMR Twin Families EBNA-1 GWAS, whether using Pearson correlation ($P = 0.752$, $r = -0.013$) or Spearman's correlation ($P = 0.453$, $r = -0.030$); indicating that MS case status itself does not predict anti-EBNA-1 titres. Below is the scatter plot for the correlation. Here the GRS_5770 refers to the constructed GRS values using 5,770 SNPs, while the Res_EBNA1 refers to the residuals of anti-EBNA-1 titres in QIMR Twin Families, as using actual titers was not appropriate due to batch effects for the measurement of anti-EBNA-1 IgG. We therefore utilized modelled residuals.

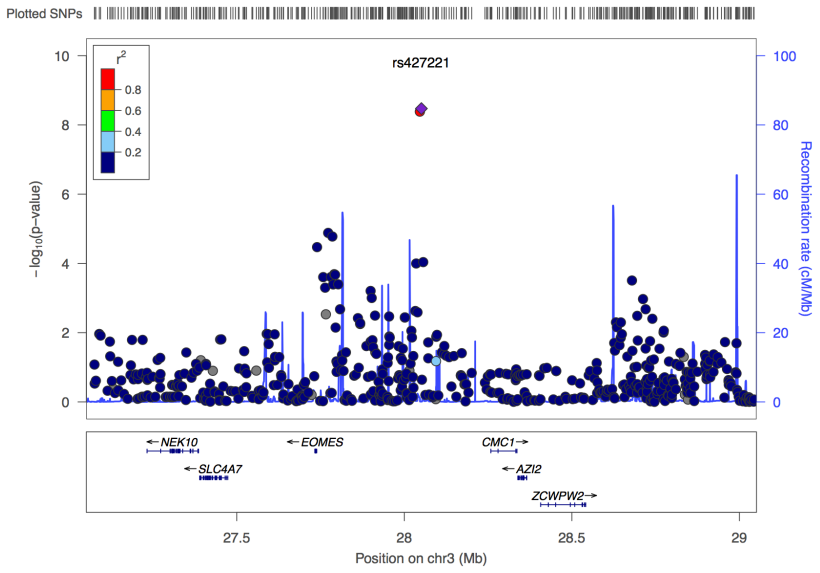


Supplementary Figure 2.9: Regional plots for the significant non-HLA loci in the joint EBNA-1 and MS meta analysis.

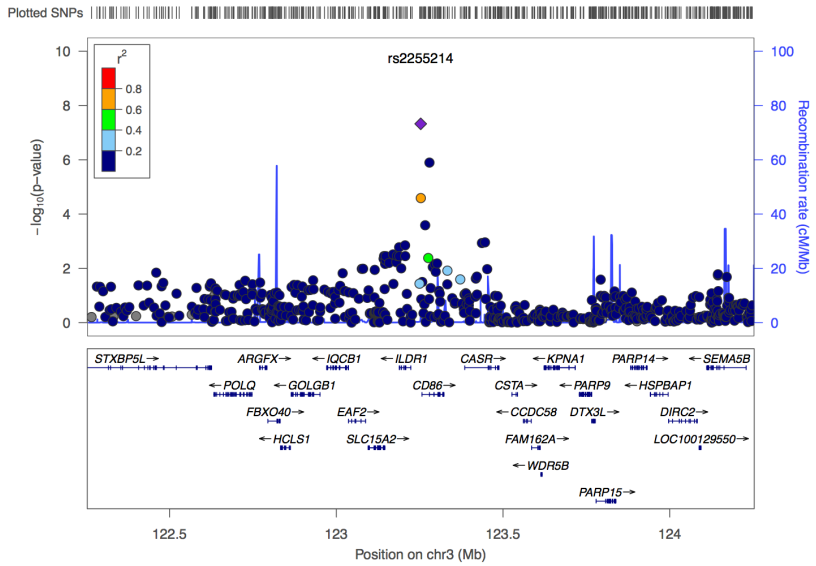
a. Regional plots for the significant loci (1p22.1). The top significant SNP is rs11808092 in the gene *EVI5*.



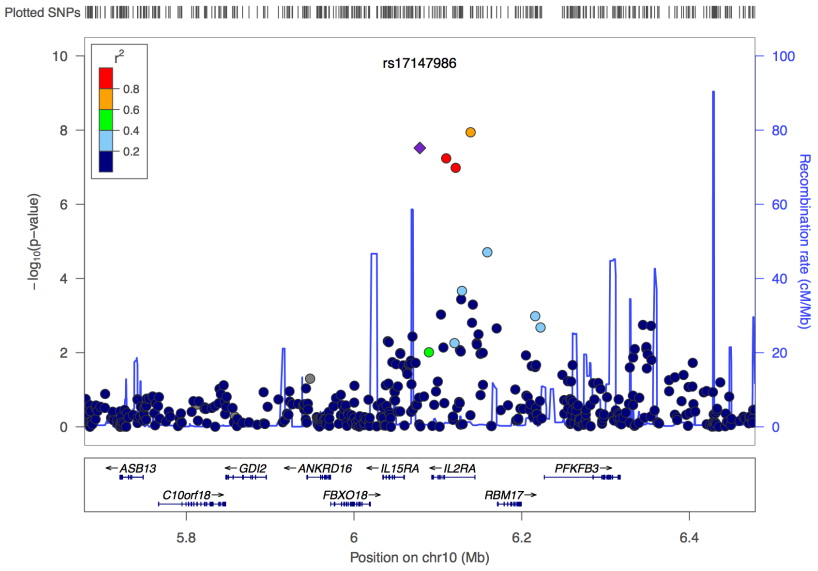
b. Regional plot for the significant loci (3p24.1). The top significant SNP is rs427221.



c. Regional plot for the significant loci (3q13.33). The top significant SNP is rs2255214.



d. Regional plot for the significant loci (10p15.1). Significant SNP rs17147986 is used as reference.



Supplementary Table 2.1:

a. Descriptive data for QIMR Twin Families (N = 3,760 individuals). MZ represents monozygotic twins. DZ represents dizygotic twins.

Groups		Sex		Age (years)	
		Female	Male	Mean	Std. Deviation
Parents	Fathers	0	661	47.54	5.22
	Mothers	725	0	45.31	4.53
Children	MZ	473	466	15.48	1.61
	DZ	465	467	15.5	1.6
	Siblings	270	233	15.26	2.87
	Total	1933	1827	26.84	15.31

b. The number of complete pairs and single twins.

Twins	Zygosity		
	MZ	DZ	Total
Single Twin	24	55	79
Twin Pair	333	563	896

Supplementary Table 2.2: Top 45 significant SNPs within the HLA region in EBNA-1 GWAS Meta-Analysis (EGMA)(effective sample size: 5,555) and related P-values for these SNPs in MS GWAS Meta-Analysis (MGMA)(effective sample size: 17,698) (fixed effect). Of these, 40 SNPs are in complete LD ($r^2 = 1$) and showed the same direction of effect; these are in weak LD ($r^2 = 0.089$) with the other 5 SNPs ($r^2 = 1$ for these five SNPs) which showed opposite direction of effect. E(A) represents the effect allele. Effect(EGMA) represents the beta value for each SNP in EGMA. P(EGMA) represents the p-value for SNPs in EGMA. P(QTFE) represent the p-value for QTFEGWAS(parents+children). P(MAFE) represents the p-value for MAFEGWAS. OR(MS) represents the OR value for each related SNP in MGMA. P(MS) represents the p-value for each related SNP in MGMA.

SNP	CHR	POS	E(A)	O(A)	Effect(EGMA)	P(EGMA)	P(QTFE)	P(MAFE)	OR(MS)	P(MS)
rs2516049	6	32678378	T	C	0.204	3.32E-20	4.11E-09	6.43E-14	1.321	4.49E-20
rs9268853	6	32537621	T	C	0.188	9.81E-18	3.30E-08	6.18E-12	1.415	4.59E-30
rs2395185	6	32541145	G	T	0.184	3.97E-17	6.95E-08	9.76E-12	1.413	1.02E-29
rs204999	6	32217957	A	G	0.180	9.78E-14	7.03E-05	1.14E-13	1.133	1.01E-04
rs3891175	6	32742445	C	T	0.194	1.41E-13	5.25E-06	2.89E-11	1.139	1.31E-04
rs3130284	6	32248465	T	C	0.186	9.15E-12	3.24E-05	5.81E-11	1.166	1.37E-05
rs3096697	6	32242488	G	A	0.186	9.68E-12	3.33E-05	6.07E-11	1.159	5.25E-04
rs2395182	6	32521295	T	G	-0.169	1.04E-11	5.22E-06	2.47E-07	0.426	3.68E-163
rs3134608	6	32225949	T	G	0.184	1.70E-11	2.17E-05	3.94E-10	1.151	7.05E-05
rs3132965	6	32254975	A	G	0.183	1.93E-11	3.35E-05	2.92E-10	1.168	1.08E-05

rs3134945	6	32254470	C	A	0.183	1.93E-11	3.24E-05	2.92E-10	1.167	1.14E-05
rs3134947	6	32253183	C	T	0.183	1.93E-11	3.24E-05	2.92E-10	1.166	1.29E-05
rs5020946	6	32558067	G	T	0.143	3.75E-11	7.68E-06	2.01E-07	1.556	7.95E-54
rs443198	6	32298384	A	G	0.141	6.37E-11	6.94E-08	1.61E-04	1.297	3.33E-17
rs3129888	6	32519704	A	G	-0.163	1.04E-10	5.22E-06	3.78E-06	0.430	1.34E-161
rs3129882	6	32517508	A	G	-0.134	1.14E-10	3.15E-07	1.42E-04	0.624	7.88E-65
rs2073045	6	32447526	G	A	0.139	1.79E-10	3.71E-05	1.48E-07	1.403	4.44E-19
rs204993	6	32263559	A	G	0.151	2.98E-10	2.18E-06	1.74E-05	1.260	2.70E-12
rs9268516	6	32487467	C	T	0.138	5.65E-10	4.27E-05	1.71E-06	1.404	5.35E-26
rs3763316	6	32484724	C	T	0.138	6.27E-10	4.27E-05	2.05E-06	1.380	3.02E-24
rs9268494	6	32483330	A	C	0.136	7.22E-10	5.65E-05	1.71E-06	1.375	6.41E-24
rs2076524	6	32478662	A	G	0.136	1.10E-09	5.65E-05	2.64E-06	1.395	6.81E-26
rs2076525	6	32478594	T	C	0.136	1.10E-09	5.65E-05	2.64E-06	1.412	3.49E-28
rs3817962	6	32476292	C	A	0.136	1.10E-09	5.65E-05	2.57E-06	1.352	2.39E-14
rs9268474	6	32465143	T	C	0.136	1.10E-09	5.92E-05	2.64E-06	1.396	6.14E-26
rs204995	6	32262263	A	G	0.159	1.17E-09	1.57E-04	6.48E-08	1.156	3.95E-05
rs9268493	6	32483308	G	A	0.134	1.51E-09	1.48E-04	6.73E-07	1.327	4.40E-20
rs2395157	6	32456123	A	G	0.134	1.65E-09	1.12E-04	1.90E-06	1.387	5.71E-25
rs204994	6	32262976	C	T	0.160	1.66E-09	1.20E-04	1.16E-07	1.157	3.15E-05
rs3817963	6	32476065	T	C	0.134	2.01E-09	5.65E-05	6.14E-06	1.395	6.55E-26
rs3763311	6	32484154	C	T	0.133	2.12E-09	4.14E-05	9.37E-06	1.357	2.87E-14
rs9268403	6	32449451	T	C	0.133	2.77E-09	1.12E-04	3.44E-06	1.389	4.96E-25
rs3130048	6	31721718	T	C	0.148	3.52E-09	2.40E-03	1.54E-10	1.182	3.02E-07
rs176095	6	32266297	A	G	0.153	5.92E-09	1.56E-04	6.74E-07	1.154	3.84E-05
rs2267647	6	33083489	G	A	-0.124	6.51E-09	8.90E-07	1.58E-03	0.920	2.70E-03
rs2076536	6	32447326	T	C	-0.131	8.04E-09	3.53E-05	2.79E-05	0.590	1.34E-49
rs6901158	6	32313920	C	T	0.188	1.42E-08	1.31E-04	3.65E-07	1.136	2.48E-03

rs9268499	6	32483673	G	A	0.124	2.52E-08	1.48E-04	2.69E-05	1.328	3.76E-20
rs1265159	6	31248026	G	A	0.142	2.89E-08	5.36E-04	9.19E-07	1.050	1.60E-01
rs652888	6	31959213	A	G	0.151	3.05E-08	1.81E-03	5.08E-09	1.195	3.03E-06
rs4248166	6	32474399	T	C	-0.160	9.55E-10	1.86E-02	9.19E-11	1.274	2.74E-11
rs2294881	6	32475582	T	C	-0.152	3.87E-09	1.86E-02	1.11E-09	1.271	4.64E-11
rs2294882	6	32475493	T	C	-0.152	3.87E-09	1.86E-02	1.11E-09	1.240	1.70E-06
rs2294884	6	32475237	T	G	-0.151	7.19E-09	1.86E-02	2.53E-09	1.246	2.73E-07
rs10947261	6	32481210	G	T	-0.177	3.90E-08	9.95E-02	2.02E-08	1.061	2.35E-01

Supplementary Table 2.3: The combined meta-analysis of EBNA-1 GWAS Meta-Analysis (EGMA) results(N = 5,555) and MS GWAS Meta-Analysis results(N = 17,698) (MGMA) (excluding the HLA region); only significant SNPs are listed (P_Meta < 5.0×10^{-8}). E(A) represents the effect allele. P_MGMA represents the p-value for SNPs in MGMA. P_EGMA represents the p-value for related SNPs in EGMA. P_Meta represents the meta-analysis results of the MGMA and EGMA. I^2 represents the percentage of total variation in SNP effects across MGMA and EGMA. Phet represents p-value for heterogeneity between EGMA and MGMA.

SNP	E(A)	O(A)	CHR	POS	P_MGMA	OR_MGMA	P_EGMA	Beta_EGMA	P_Meta	I^2	Phet
rs11808092	A	C	1	92845816	2.94×10^{-06}	1.149	1.25×10^{-03}	0.025	1.71×10^{-08}	0	0.605
rs669607	A	C	3	28046448	2.10×10^{-08}	0.854	0.0484	-0.041	4.12×10^{-09}	0	0.659
rs427221	T	C	3	28050989	2.28×10^{-08}	0.855	0.0382	-0.043	3.72×10^{-10}	0	0.738
rs2255214	T	G	3	123253229	4.95×10^{-07}	0.872	0.0300	-0.047	4.73×10^{-08}	36.4	0.209
rs17147986	C	A	10	6078484	2.57×10^{-08}	1.274	0.1910	0.043	3.04×10^{-08}	0	0.812
rs12722561	C	T	10	6109899	2.20×10^{-08}	1.240	0.3110	0.033	5.78×10^{-08}	0	0.553
rs2104286	T	C	10	6139051	3.52×10^{-10}	1.215	0.7450	0.009	1.15×10^{-08}	57.8	0.124

Supplementary Table 2.4: Interaction analysis (p-value) results for two paired SNPs (MS GWAS obtained from dbGaP and ANZgene (cases = 3,252; controls = 5,725)). $P < 0.005$ were defined as significant after Bonferroni correction.

SNP	rs11808092	rs427221	rs2255214	rs2516049	rs12722561
rs11808092	NA	0.656	0.389	0.006	0.311
rs427221	0.656	NA	0.917	0.416	0.671
rs2255214	0.389	0.917	NA	0.177	0.158
rs2516049	0.006	0.416	0.177	NA	0.198
rs12722561	0.311	0.671	0.158	0.198	NA

Supplementary Table 2.5 Detailed interaction results for rs2616049 (between *HLA-DRB1* and *HLA-DQA1*) and rs11808092 (*EVI5*) at each genotype level. The reference genotype is C/C-C/C

Genotype	Frequency (%)	OR (95% CI)	P-value
C/C-C/C	400 (4.5)	1.00 [Reference]	
C/C-C/A	318 (3.5)	0.83 (0.56, 1.20)	*0.47*
C/C-A/A	51 (0.6)	0.88 (0.42, 1.83)	*0.90*
C/T-C/C	2,046 (22.8)	1.63 (1.26, 2.13)	*0.11*
C/T-C/A	1,442 (16.1)	1.76 (1.34, 2.30)	*0.96*
C/T-A/A	238 (2.7)	**3.05 (2.10, 4.41)**	$\boldsymbol{6.73\times10^{-5}}$
T/T-C/C	2,350 (26.2)	**2.23 (1.71, 2.88)**	***0.04***
T/T-C/A	1,826 (20.4)	**2.68 (2.06,3.45)**	***0.01***
T/T-A/A	301 (3.4)	2.25 (1.59, 3.18)	*0.59*
Trend:			***P = 0.006***

Genotype presented as rs2616049 genotype/rs11808092 genotype.

Appendix 2.A: Publication of "Chapter 2: Genetic loci for Epstein-Barr Virus nuclear antigen-1 are associated with risk of multiple sclerosis"

Zhou Y, Zhu G, Charlesworth JC, et al. Genetic loci for Epstein-Barr virus nuclear antigen-1 are associated with risk of multiple sclerosis. *Mult Scler* 2016.

Chapter 3: Genetic variation in PBMC-produced IFN-γ and TNF-α association with relapse in multiple sclerosis

3.1 Preface

The preceding chapter reported a cross-sectional design approach in a large sample size study that was aimed at investigating the association of genetic variation with MS development and the role that the genetic variation plays. This strategy is widely used in GWAS because it is relatively easy to obtain large sample sizes. However, there is currently very little research based on longitudinally designed studies for finding genetic variation in longitudinally monitored patients to predict the clinical course of MS. In this chapter and the next two chapters, I will present my work focusing on this area. In this chapter, I will assess whether SNPs within a number of genes encoding cytokines and their receptors, which have been strongly associated with MS inflammatory relapses, significantly modulate the associations of the pro-inflammatory cytokines TNF-α and IFN-γ with MS relapse, thus providing information about the effects of these cytokines and the potential role of their respective genetic loci in the MS clinical course. This chapter was published in the Journal of Neurological Sciences[1] (Appendix 3.A).

3.2 Introduction

In our previous work, we found that stimulated peripheral blood mononuclear cell (PBMC)-produced TNF-α was associated with a lower risk of relapse, while stimulated PBMC-produced IFN-γ was significantly associated with an increased risk

of relapse. We further found that one MS-related SNP, rs1800693, within the TNF-α-receptor gene significantly modified the TNF-α association with relapse, such that the secreted form blocked the protective effect of TNF-α[2].

In this paper, we provide further genetic context by assessing whether SNPs within a greater number of genes for cytokines and their receptors significantly modulate the associations of TNF-α and IFN-γ with relapse, thus providing information about the effects of these cytokines and the potential role of their respective loci in MS.

3.3 Methods

3.3.1 Study design

The Southern Tasmanian Multiple Sclerosis Longitudinal Study (MSL) followed a cohort of 203 persons with clinically definite MS living in southern Tasmania, Australia over 2002 – 2005[3]. As the primary clinical outcome was the occurrence of relapse during the study, analysis was restricted to the 119 persons with relapsing-remitting MS (Table 3.1) followed beyond one review with data on all cytokines (IL4, IL10, IFN-γ and TNF-α), of which, 91 participants had data on genotype.

At each biannual review participants were asked about relevant environmental and behavioural factors, including medications, smoking, physical activity, time in the sun and diet.

Ethics approval was obtained from the Southern Tasmania Human Research Ethics Committee. All participants provided informed consent.

3.3.2 Measurement of relapses

A relapse was defined according to the 2001 McDonald Criteria[4] as the acute or subacute appearance or reappearance of a neurological abnormality (lasting at least 24 hours), immediately preceded by a stable, improving, or slowly progressive neurological state for 30 days, in the absence of fever, known infection, concurrent steroid withdrawal, or externally derived increases in body temperature. Relapses were reported in real time by phone or at biannual review, and all reports validated by the study neurologist. While MRI was a component of neurologist diagnosis of MS conversion, it was not a component of our relapse allocation which was just symptom based. Since there weren't systematic MRIs so the possibility of persons having subclinical exacerbations which might be detectable on MRI could not be systematically measured across all participants. Thus our definition of relapse was just clinical.

3.3.3 Biological samples

At each biannual review, blood samples were obtained by standard venepuncture and processed to separate serum and buffy coat on the day by study personnel. All reviewers were fact to face. Relapse reports were done by phone but these did not entail a blood sample. For two of these reviews, Summer 2004 and Winter 2004, buffy coats were processed to separate out PBMCs: collected blood was layered over histopaque (Sigma), spun at 400 g for 30 minutes, and the interface containing the PBMC was removed and washed twice in incomplete RPMI-1640 (JRH Biosciences, Melbourne, Australia). The cells were then resuspended in a 10% DMSO/RPMI solution and stored at - 80°C until required. To avoid a storage effect, as soon as a

season's collection was completed, the cells were thawed and counted using a haemocytometer. Cell viability was ascertained by Trypan blue exclusion, and any samples with viability below 90% were discarded. Cells were then washed twice in incomplete RPMI and resuspended in complete RPMI containing 2 mM L-glutamine (Sigma, Castle Hill, Australia), 100U Penicillin, 100U Gentamicin (Pharmacia, Murarrie, Australia) and 10% FCS (JRH Biosciences) at 1×10^6cells/ml in 24 well cell culture plates. Five μg/ml phytohaemaglutinin (PHA) (Sigma) was added to each sample, and the samples were incubated for 24 h at 37°C, 8% CO2. Supernatant was then stored at -80°C for the duration of the study, for analysis by ELISA thereafter.

3.3.4 ELISA protocol

To avoid a batching effect, the ELISA was performed at a single time by a single operator. Samples for IL-4 and IL-10 analysis were assayed undiluted, and those for IFN-γ and TNF-α analyses were diluted 1:4 with Assay Diluent (BD Biosciences, San Diego). ELISAs were performed using BD OptEIA ELISA kits and reagents, as per the manufacturer's instructions. To ensure consistency, each plate (Nunc, Roskilde, Denmark) contained matched MS summer/winter samples, control samples and standards. Preliminary analyses showed that storage of supernatants for approximately 12 months had no effect on cytokine levels.

3.3.5 SNP Identification & Genotyping

SNPs (N=361) within a window of 10kb around all the cytokine genes and their receptors genes (total N=83, Supplementary Table 3.1) were selected, of which, 3 (*IL9R*, *IL3RA* and *IL13RA1*) are within the X chromosome, 7 genes (*IFNA1*, *IL2*,

IL28B, *LI4*, *IL8*, *IL8RA* and *TXLNA*) with small window having no SNPs available in the datasets, leaving 73 genes used for analysis. Different gene boundaries (5kb,10kb,20kb and 50kb) have different advantages and limitations[5]. The use of conservative window (10kb) in this paper is partly as a function of reducing multiple comparison issues.

These SNPs were directly genotyped in 189 of our samples as either part of the ANZgene MS GWAS or by an additional GWAS, extensively described in our previous study[6]. Briefly, 164 MS cases were genotyped as a part of ANZgene MS GWAS and additionally 29 MS cases were newly genotyped using the Illumina HumanOmniExpress-12v1_A array. After quality control, 4 duplicated samples were removed. All the samples were previously identified as being from individuals of European descent[7].

3.3.6 Statistical analysis

As in the previous work[2], the associations of cytokines with time to relapse were evaluated in a multivariable model wherein IFN-γ, TNF-α and IL-10 were mutually adjusted, while a categorical of IL-4 was stratified upon (allowing the baseline hazard to vary by level of the IL-4 categorical) due to its violation of the proportional hazards assumption. Stratification in this context means a statistical procedure whereby the baseline hazard was allowed to "float" by level of the variable stratified on, in this case the 4 levels of the IL-4 categorical.

Predictors of PBMC cytokines were evaluated using multilevel mixed-effects linear regression. Transformation was applied as required to satisfy homoscedasticity; however all coefficients are reported on the scale of the original value.

All covariates satisfied the proportional hazards assumption, except IL-4, so all models were stratified by a four-level categorical rendition of IL-4 (11.5 – 52.9, >52.9 – 86.8, >86.8 – 127.6, >127.6 – 511.9).

Interaction of SNPs and cytokines with time to relapse was evaluated by Cox proportional hazards for repeated events, as described previously[8], whereby multiple relapses by the same persons are treated as independent observations but accounted for at the intra-individual level, and the time until a prior event does not influence the composition of the risk set for a subsequent event. Where interaction was assessed, a product term of the SNP and cytokine was generated and included in the multivariable model, the significance of this term defining significance of statistical interaction.

We defined $p<1.39\times10\text{-}4$ as a threshold of significance after adjusting for multiple comparison with Bonferroni correction. All analyses were performed using STATA/SE for Windows (Version 12.1; StataCorp LP College Station, TX USA).

3.4 Results

As discussed previously, our cohort was majority female, of middle age and low disability, and a minority had any relapses, both during the study and during the analysis period here, reflecting their largely being on immunomodulatory therapy (Table 3.1).

Table 3.1 Cohort characteristics of sample analysed.

	n (%)
All persons	119
Sex	
Male	31 (26.1)
Female	88 (74.0)
Age at study entry (years)	
21 – 39	30 (25.2)
>39 – 45	28 (23.5)
>45 – 52	31 (26.1)
>52 – 77	30 (25.2)
Progression to SPMS during study	
No	105 (88.2)
Yes	14 (11.8)
Relapse during study?	
No	
Yes	59 (49.6)
Relapse during analysis period?	
No	
Yes	33 (27.7)
Any immunomodulatory therapy during study?	
No	
Yes	98 (82.4)
MS duration from symptom onset (years)	10 (5 – 17)
EDSS	2.5 (2 – 4)
Number of relapses during analysis period	0 (0 – 1)

SPMS = Secondary-progressive multiple sclerosis; EDSS = Expanded Disability Status Scale.
The use of medication includes Betaferon, Avonex, Rebif and Copaxone.

3.4.1 Cytokine SNPs to evaluate associations with TNF-α and IFN-γ production

After adjusting for multiple testing using the Bonferroni correction method ($P<1.39\times10^{-4}$ was defined as significant), we did not observe any cytokine-related SNPs which significantly impacted on TNF-α and IFN-γ production (data not shown).

3.4.2 Differential associations of TNF-α and IFN-γ with time to relapse by SNP genotype

We found one SNP, rs3218295, located within the gene *IL2RB,* which showed a significant interaction with TNF-α as a predictor of time to relapse ($p_{interaction}=5.04\times10^{-5}$). Carriers of the GG genotype showed a significant protective effect of TNF-α on relapse ($p=1.46x10^{-5}$), while carriers of the GA or AA genotypes showed a significantly increased risk of relapse from increased TNF-α (p=0.001) (Table 3.2). Additionally, the SNP rs522807, located within the 3' region of the gene *TNFRSF1B,* interacted with TNF-α's reduction in the risk of relapse ($p_{interaction}=3.83\times10^{-4}$), such that carriers of the minor A allele showed a much greater protective effect of TNF-α on relapse risk than those homozygous for the CC genotype, who had an association no different from the parent cohort.

Table 3.2 SNP for an interactive association with TNF-α as a predictor of time to relapse.

SNP	Type	Related Genes	Genotype	# of Relapses	HR (95%CI)	Significance
Aggregate					0.48 (0.18,1.29)	*p=0.150*
rs522807 (A/C)	3' region, intronic	*TNFRSF1B*	CC	35	0.47 (0.15, 1.43)	*p=0.180*
	TFBS		CA+AA	3	**0.05 (0.01, 0.32)**	***p=0.001***
						$p_{interaction}=3.83x10^{-4}$
rs3218295 (A/G)	Within gene, intronic	*IL2RB*	GG	30	**0.18 (0.08, 0.39)**	$p=1.46x10^{-5}$
			GA+AA	8	**7.41 (2.20, 24.95)**	***p=0.001***
						$p_{interaction}=5.04x10^{-5}$

All analyses mutually adjusted for cytokines (IFN-γ, TNF-α, and IL-10), and for age, sex, time in sun in preceding 2/52 (time in the sun in the preceding two weeks before review), ever smoked, IFN-beta medication use (sensitivity analysis showed similar effect as adjusting for immunomodulatory therapy), and stratified by IL4 quartile categorical. Exposure to EBV was not a significant covariate in the analysis. Results are presented for 100-unit increments of the cytokine.

$P_{interaction}$ <0.0001 was defined as significant($P_{interaction}$ represents the P value for interaction)

Abbreviations: TFBS: transcriptional factor binding sites.

Table 3.3 SNP for an interactive association with IFN-γ as a predictor of time to relapse.

SNP	Type	Related Genes	Genotype	# of Relapses	HR (95%CI)	Significance
Aggregate					**1.75 (1.04, 2.94)**	***p=0.035***
rs522807 (A/C)	3' region, intronic	*TNFRSF1B*	CC	35	**2.51 (1.46, 4.30)**	***p=0.001***
	TFBS		CA+AA	3	0.56 (0.25, 1.27)	*p=0.160*
						$p_{interaction}=8.21x10^{-5}$
rs25879 (G/A)	5' region, intronic	*IL3/CSF2*	AA	29	**2.65 (1.48, 4.74)**	***p=0.001***
	TFBS		AG+GG	9	0.50 (0.21, 1.22)	*p=0.130*
						$p_{interaction}=1.7x10^{-5}$

All analyses mutually adjusted for cytokines (IFN-γ, TNF-α, and IL-10), and for age, sex, time in sun in preceding 2/52(time in the sun in the preceding two weeks before review), ever smoked, IFN-beta medication use (sensitivity analysis showed similar effect as adjusting for immunomodulatory therapy), and stratified by IL4 quartile categorical. Results are presented for 100-unit increments of the cytokine.

$P_{interaction}$ <0.0001 was defined as significant($P_{interaction}$ represents the P value for interaction). Exposure to EBV was not a significant covariate in the analysis

Abbreviations: TFBS: transcriptional factor binding sites.

We also found two SNPs showing interactive associations with IFN-γ as a predictor of time to relapse (Table 3.3). The rs522807 SNP, located within the 3' region of the gene *TNFRSF1B*, showed a significant interaction with IFN-γ ($p_{interaction}=8.21x10^{-5}$), such that carriers of the CC genotype had an enhanced risk of relapse from increased IFN-γ, while carriers of the CA or AA genotypes showed no association between IFN-γ and relapse. Likewise the rs25879 SNP, located within the 5' region of the gene of *IL3*, interacted significantly with IFN-γ ($p_{interaction}=1.70x10^{-5}$), showing a significantly increased association of IFN-γ with relapse among carriers of the AA genotype, while carriers of the AG or GG genotypes showed no association between IFN-γ and relapse. Both SNPs are within the transcriptional factor binding sites of their respective genes.

3.5 Discussion

In this work, we systematically examined the SNPs within a large number of genes for cytokines and their receptors to assess whether there was an interaction with the previously demonstrated associations[2] of TNF-α and IFN-γ with relapse. Individuals of GG genotype of rs3218295 (within the gene *IL2RB*) demonstrated a significant protective effect of TNF-α on relapse while those of GA/AA genotype did not ($P_{interaction} = 5.04\times10^{-5}$). Carriers of CC genotype of rs522807 (3' region of *TNFRSF1B*) and the AA genotype of rs25879 (5'region of *IL3*) showed a strong association between IFN-γ and increased relapse risk ($P_{interaction} =8.21\times10^{-5}$ and 1.70×10^{-5} respectively). That we have here shown allelic variants in a SNP in the TNF-α receptor gene modulates both the associations of TNF-α and IFN-γ, with carriers of the A allele showing a potentiated protective effect against relapse from

increased TNF-alpha, and an abrogation of the increased relapse hazard from increased IFN-gamma, is supportive of a true effect for this SNP. The finding of significant interactions with SNPs in the IL-2 receptor and IL-3 genes with TNF-α and IFN-γ, respectively, are also interesting and reflect the complex interplay between the various cytokines and cytokine receptors of the immune system.

Despite multiple MS risk GWAS[9,10], no susceptibility loci in or near the gene of *IL2RB*[9] or *IL3* have previously been demonstrated, though certainly TNF-α receptor is now a well-recognised MS susceptibility locus[11]. That these loci have not been identified in MS GWAS is not unexpected, because GWAS are focused more on cross-sectional data and are limited by a greater multiple testing burden. This reflects the utility in examining SNPs in parameters relevant to the already demonstrated cytokine risk factors, because by casting the net in this fashion and using a systems biology approach, we have identified results that may uncover modulators of these biological risk factors. Similar approaches have been taken with regard to vitamin D/UV[12-14], childhood infections[15] and EBV[16] that have either uncovered novel mechanisms of already known risk alleles for MS [12-16] or uncovered novel loci that modulate the course of MS [17].

Our findings of effects for the three SNPs presented here are not wholly novel in terms of relevance to neurological function. The impact of *TNFRSF1B* SNPs on MS has been well-demonstrated[11]. Our other SNPs have additional evidence in support of them being biologically plausible candidates. A locus near *IL3* (rs31480) is in high linkage disequilibrium with our significant SNP (rs25879, r2>0.8). This SNP could influence the binding affinity of the transcription factor SP1 and thus impact upon the expression of *IL3*. Others have shown that the level of IL3 promoted the proliferation

and survival of neural progenitors [18]. Consistent with this finding, our major allele genotype AA of rs25879 and the major allele genotype CC of rs31480 showed the same effect direction. That is, AA of rs25879 interacted significantly with IFN-γ to increase relapse risk, while CC of rs31480 in their findings was associated with lower brain volume and lower IL3 expression which would decrease neural progenitors' survival and proliferation. These results suggest a possible mechanism by which IL3 might interact with IFN-γ, potentially regulating brain processes and affecting susceptibility to relapse.

IL2RB encodes the beta chain of the IL-2 receptor, expressed on T and B-cells, monocytes, neutrophils and dendritic cells. In mice, the blocked expression of *IL2RB* manifested in the dysregulation of T and B-cell activation and behaviour, as well as the loss of thymocytes, leading to death by 12 weeks of age[19]. Thus, IL-2R-beta appears to be required for proper control and function of immune cell activation, reducing the potential for inappropriate immune activation[19]. IL-2R-beta can bind to IL-15, whose three-dimensional structure is similar to that of IL-2[20]. IL-15 plays an important role in the maintenance and activation of CD8 T lymphocytes, a prominent lymphocyte population found in MS lesions[21]. In MS, IL-15 serum levels were higher in patients with relapse compared with patients in stable periods of the disease ($p=0.001$)[22]. Although a role of *IL2RB* in susceptibility to other inflammatory diseases like rheumatoid arthritis[23] and asthma[24] has been found in some studies, its role in MS remains unknown, so further studies are required to better understand its role in disease.

This study benefits from its prospective cohort design and real-time reporting of relapses. The current analysis is limited to some extent in its being conducted on only

a subset of the MSL study cohort, but nonetheless provides valuable information on the nature of the immune response in MS clinical course and the potential genetic impacts thereon. The biannually obtained PBMCs do reasonably represent the immune state between relapses, allowing these to be a predictor of subsequent relapse risk. There is variability over time due to changing season, changes in medication and infection state, among others, which can impact on the responsiveness of immune cells. Our use of a standardised and appreciable stimulus to induce cytokine production[25] should allow a comparable cytokine production between persons that can overcome this micro-variation over time to give a representation of the general immune state of the individual.

In conclusion, our results provide insight into the complex interaction of cytokines that affect the risk of relapse in MS. Due to the complexity of the immune system, having measurements of other cytokines would be relevant and future studies seeking to confirm these findings are needed in order to establish them for future use in clinical practice. These findings may provide some clues to better understand the immunopathology of MS and may suggest possible points of intervention in moderating clinical course. They also highlight the potential of more targeted future therapeutics for MS that take into account a person's genotype.

3.6 Summary

Background: Alterations in peripheral blood mononuclear cells (PBMC) cytokine production have been found in multiple sclerosis (MS) compared to healthy controls. We have previously found that stimulated PBMC-produced TNF-α and IFN-γ

modulated MS relapse risk, such that raised TNF-α was protective, while raised IFN-γ increased relapse risk.

Objective: To assess whether SNPs within genes for relevant cytokines and their receptors modulate the associations of TNF-α and IFN-γ with relapse, thus providing additional information about these cytokine effects and the roles of these genes in MS.

Methods: Prospective cohort of 91 participants with relapsing-remitting MS and cytokine and genotype data. SNPs (N=361) within a window of 10kb around each cytokine/cytokine receptor gene (N=83) selected for analysis. Predictors of PBMC cytokines evaluated by multilevel mixed-effects linear regression. Predictors of relapse evaluated by Cox proportional hazards regression. Bonferroni correction used to adjust for multiple testing; thus P<1.39x10-4 defined as significant.

Results: Individuals of GG genotype of rs3218295 (within the gene IL2RB) demonstrated a significant protective effect of TNF-α on relapse while those of GA/AA genotype showed a significant positive association (Pinteraction=5.04×10-5). Carriers of CC genotype of rs522807 (3' region of TNFRSF1B) and the AA genotype of rs25879 (5'region of IL3) showed a strong association between IFN-γ and increased relapse risk (Pinteraction=8.21×10-5 and 1.70×10-5, respectively).

Conclusions: Our results show novel modulation of TNF-α and IFN-γ associations with relapse by SNPs in major cytokines. These findings suggest the potential for these genes and/or their products as potential therapeutic targets in MS.

3.7 References

1. Zhou Y, Taylor B, van der Mei I, et al. Genetic variation in PBMC-produced IFN-gamma and TNF-alpha associations with relapse in multiple sclerosis. *J. Neurol. Sci.* 2015;349(1-2):40-44.

2. Simpson S, Jr., Stewart N, van der Mei I, et al. Stimulated PBMC-produced IFN-gamma and TNF-alpha are associated with altered relapse risk in multiple sclerosis: results from a prospective cohort study. *J Neurol Neurosurg Psychiatry.* 2014.

3. Pittas F, Ponsonby AL, van der Mei IA, et al. Smoking is associated with progressive disease course and increased progression in clinical disability in a prospective cohort of people with multiple sclerosis. *J. Neurol.* 2009;256(4):577-585.

4. McDonald WI, Compston A, Edan G, et al. Recommended diagnostic criteria for multiple sclerosis: guidelines from the International Panel on the diagnosis of multiple sclerosis. *Ann Neurol.* 2001;50(1):121-127.

5. Lin R, Taylor BV, Simpson S, Jr., et al. Novel modulating effects of PKC family genes on the relationship between serum vitamin D and relapse in multiple sclerosis. *Journal of neurology, neurosurgery, and psychiatry.* 2014;85(4):399-404.

6. Australia, New Zealand Multiple Sclerosis Genetics C. Genome-wide association study identifies new multiple sclerosis susceptibility loci on chromosomes 12 and 20. *Nat. Genet.* 2009;41(7):824-828.

7. Simpson S, Jr., Taylor B, Blizzard L, et al. Higher 25-hydroxyvitamin D is associated with lower relapse risk in multiple sclerosis. *Ann. Neurol.* 2010;68(2):193-203.

8. Cavanillas ML, Alcina A, Nunez C, et al. Polymorphisms in the IL2, IL2RA and IL2RB genes in multiple sclerosis risk. *Eur. J. Hum. Genet.* 2010;18(7):794-799.

9. International Multiple Sclerosis Genetics C, Wellcome Trust Case Control C, Sawcer S, et al. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature.* 2011;476(7359):214-219.

10. Gregory AP, Dendrou CA, Attfield KE, et al. TNF receptor 1 genetic risk mirrors outcome of anti-TNF therapy in multiple sclerosis. *Nature.* 2012;488(7412):508-511.

11. Ramagopalan SV, Hanwell HE, Giovannoni G, et al. Vitamin D-dependent rickets, HLA-DRB1, and the risk of multiple sclerosis. *Archives of Neurology.* 2010;67(8):1034-1035.

12. Lin R, Taylor BV, Simpson S, Jr., et al. Association between multiple sclerosis risk-associated SNPs and relapse and disability - a prospective cohort study. *Multiple sclerosis (Houndmills, Basingstoke, England).* 2013;20(3):313-321.

13. Ramagopalan SV, Link J, Byrnes JK, et al. HLA-DRB1 and month of birth in multiple sclerosis. *Neurology.* 2009;73(24):2107-2111.

14. Ramagopalan SV, Dyment DA, Giovannoni G, Sadovnick AD, Ebers GC. HLA-DRB1*15, low infant sibling exposure, and multiple sclerosis gene-environment interaction. *Ann Neurol.* 2010;67(5):694-695.

15. Ramagopalan SV, Sadovnick AD, Ebers GC, Giovannoni G. Effects of infectious mononucleosis and HLA-DRB1*15 in multiple sclerosis. *Multiple sclerosis.* 2010;16(1):127-128.

16. Lin R, Taylor BV, Simpson S, Jr., et al. Novel modulating effects of PKC family genes on the relationship between serum vitamin D and relapse in multiple sclerosis. *J Neurol Neurosurg Psychiatry.* 2013;85(4):399-404.

17. Luo XJ, Li M, Huang L, et al. The interleukin 3 gene (IL3) contributes to human brain volume variation by regulating proliferation and survival of neural progenitors. *PLoS One.* 2012;7(11):e50375.

18. Suzuki H, Kundig TM, Furlonger C, et al. Deregulated T cell activation and autoimmunity in mice lacking interleukin-2 receptor beta. *Science.* 1995;268(5216):1472-1476.

19. Wuest SC, Edwan JH, Martin JF, et al. A role for interleukin-2 trans-presentation in dendritic cell-mediated T cell activation in humans, as revealed by daclizumab therapy. *Nature medicine.* 2011;17(5):604-609.

20. Saikali P, Antel JP, Pittet CL, Newcombe J, Arbour N. Contribution of astrocyte-derived IL-15 to CD8 T cell effector functions in multiple sclerosis. *Journal of immunology.* 2010;185(10):5693-5703.

21. Rentzos M, Nikolaou C, Rombos A, et al. Circulating interleukin-15 and RANTES chemokine in MS patients: effect of treatment with methylprednisolone in patients with relapse. *Neurological research.* 2010;32(7):684-689.

22. Diogo D, Kurreeman F, Stahl EA, et al. Rare, low-frequency, and common variants in the protein-coding sequence of biological candidate genes from

GWASs contribute to risk of rheumatoid arthritis. *Am. J. Hum. Genet.* 2013;92(1):15-27.

23. Akhabir L, Sandford AJ. Genome-wide association studies for discovery of genes involved in asthma. *Respirology.* 2011;16(3):396-406.

24. Stewart N, Taylor B, Ponsonby AL, et al. The effect of season on cytokine expression in multiple sclerosis and healthy subjects. *J. Neuroimmunol.* 2007;188(1-2):181-186.

Supplementary Table 3.1: List of genes selected for analysis. Here the genomic location for the genes were based on hg18. Chr refers to chromosome

Gene	Start	End	Chr	Gene	Start	End	Chr
TNFRSF1B	12149646	12191863	1	IL6	22733342	22738144	7
IL22RA1	24318847	24342197	1	IL7	79807559	79880312	8
TXLNA	32417931	32436472	1	IL33	6231677	6247981	9
IL23R	67404756	67498249	1	IFNB1	21067103	21067942	9
IL12RB2	67545634	67635170	1	IFNA1	21430439	21431314	9
IL6R	152644292	152706811	1	IL11RA	34643931	34651883	9
IL10	205007570	205012461	1	TGFBR1	100907232	100956294	9
IL19	205038837	205082947	1	IL15RA	6034339	6060147	10
IL20	205105776	205109190	1	IL2RA	6093511	6144277	10
IL24	205137411	205144106	1	IL18BP	71387605	71391221	11
TGFB2	216586490	216681595	1	IL18	111519185	111540049	11
IL1R1	102136833	102162765	2	IL10RA	117362318	117377403	11
IL18RAP	102401685	102435456	2	TNFR1	6308183	6321521	12
IL1A	113247962	113259441	2	CD4	6768911	6800236	12
IL1B	113303807	113310826	2	IL23A	55018929	55020460	12
IL36A	113479920	113482092	2	IFNG	66834816	66839787	12
IL36B	113496139	113526915	2	IL26	66881395	66905837	12
IL1RN	113591940	113608063	2	IL22	66928291	66933547	12
IL8RB	218698990	218710219	2	IL31	121222529	121224698	12
IL8RA	218735812	218739960	2	IL25	22911857	22915451	14
IL5RA	3086420	3127030	3	TGFB3	75494194	75517241	14
IL17RC	9933781	9950313	3	IL16	79276273	79392156	15
IL17RB	53855616	53874866	3	IL32	3055313	3059668	16
IL12A	161189322	161196499	3	IL4R	27232751	27283599	16
IL8	74825138	74828296	4	IL21R	27321223	27369615	16
IL2	123592074	123597099	4	IL17C	87232501	87234382	16
IL21	123739271	123761661	4	EBI3	4180539	4188524	19
IL15	142777203	142874061	4	IL27RA	14003261	14025025	19
IL7R	35892747	35912680	5	IL12RB1	18031370	18058696	19
IL6ST	55272450	55326519	5	IL28B	44426111	44427450	19
IL3	131424245	131426794	5	IL28A	44450996	44452571	19
CSF2	131437384	131439758	5	IL29	44478804	44481151	19
IL5	131905034	131907112	5	TGFB1	46528490	46551655	19
IL13	132021763	132024699	5	IL11	60567568	60573625	19
IL4	132037271	132046266	5	IFNAR2	33524100	33558696	21
IL9	135255833	135259414	5	IFNAR1	33619083	33653998	21
IL17B	148734022	148739030	5	IL17RA	15945848	15971404	22
IL12B	158674368	158690058	5	IL2RB	35851823	35875907	22
LTA	31648071	31650076	6	IL3RA	1415508	1461581	X|Y
TNF	31651328	31654090	6	IL13RA1	117745586	117812523	X
LTB	31656313	31658180	6	IL9R	154880439	154893675	X|Y
IL17A	52159143	52163394	6				

Appendix 3.A: Publication of "Genetic variation in PBMC-produced IFN-γ and TNF-α associations with relapse in multiple sclerosis "

Zhou Y, Taylor B, van der Mei I, et al. Genetic variation in PBMC-produced IFN-gamma and TNF-alpha associations with relapse in multiple sclerosis. *Journal of the neurological sciences* 2015; **349**(1-2): 40-4.

Chapter 4: Variation within the myelin basic protein gene predicts disease course in multiple sclerosis

4.1 Preface

The previous chapter provided evidence that genetic variations in major cytokine genes modulate the associations of the cytokines TNF-α and IFN-γ with relapse using a prospective cohort of 91 participants followed for 3 years. In this chapter, we will assess whether genetic variation in myelin basic protein may directly or by interaction with HHV6 or EBV determine clinical outcomes (conversion to MS after a first demyelinating event, relapse rate and disability) using the prospectively followed Ausimmune/Auslong Longitudinal Study, an ongoing > 10 -year prospective cohort study following a population of 282 patients who were initially identified as having a first demyelinating event (FDE) suggestive of, but not diagnostic for, MS.

4.2 Introduction

Myelin basic protein (MBP) is a major component of the myelin sheath and is believed to play an important role in the process of myelination in the CNS[1]. In patients presenting with a first demyelinating event (FDE), the presence of serum antibodies against MBP significantly predicts conversion to clinically definite multiple sclerosis (CDMS)[2]. Despite its potential role in MS, none of the large genome-wide association studies (GWAS)[3,4] have found any variants in MBP that predict MS risk.. In fact no MS GWAS has been able to establish any markers that predict severity or clinical course. Therefore, candidate gene approaches with a priori

hypotheses can be used to avoid the burden of correcting for the large number of tests associated with GWAS studies.

Emerging evidence suggests that autoimmune responses targeting the myelin sheath indicates the existence of myelin-reactive T-cells, which can recognise MBP peptides, trigger an immune response against MBP and thus affect the integrity of the myelin sheath[5]. The molecular mimicry hypothesis suggests that these myelin-reactive T-cells can be activated because certain viral agents, particularly human herpesvirus-6 (HHV6) and Epstein-Barr virus (EBV), share similar antigenic profiles with MBP, resulting in T-cells cross-reacting to both virus and MBP[6,7].

Given the high prevalence of prior exposure and seropositivity (>90%) against HHV6[8] and EBV[9] in the general population, and the existence of myelin-reactive T-cells in healthy individuals[10], we have therefore hypothesised that by using a biologically plausible candidate gene approach; that genetic variations in MBP may directly, or by interaction with HHV6 or EBV, determine clinical outcomes (conversion to MS after a first demyelinating event, relapse rate and disability). We have therefore studied this a priori hypothesis in a well-characterised cohort who had an FDE at baseline, and who had both genetic data and EBV and HHV6 antibody serology.

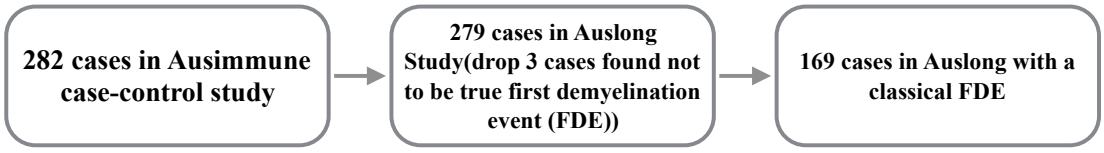
4.3 Material and Methods

4.3.1 Study design

The Ausimmune Longitudinal (AusLong) Study, which built upon the original Ausimmune case-control study, seeks to elucidate environmental, genetic and

personal risk factors for the onset and early progression of MS. This study has followed 169 cases with a classical first demyelination event (FDE)[11](Figure 4.1). The present analysis is for the period from first recorded symptom onset, to the 5-year review, as this is the most recent face-to-face review which all currently enrolled participants have completed.

Figure 4.1: Flow chart of the Auslong Study



The Ausimmune Study and AusLong Study were approved by nine regional Human Research Ethics Committees (Health & Medical Research Ethics Committee, University of Tasmania, Project No. H0010499; Barwon Health Human Research Ethics Committee (HREC), Project No. 09/24, Project No. 03/46; Hunter New England HREC, John Hunter Hospital, Project No.HREC/09/HNE/139, Project No. 09/04/15/5.04; Griffith University HREC, Project No: MED/02/10/HREC; Royal Brisbane and Women's Hospital, Queensland Health, Project No: HREC/09/QRBW/299; Queensland Institute for Medical Research, Project No. QIMR-HREC /P1252; HREC of Northern Territory Department of Health & Menzies School of Health Research, Project No. HREC 2011-1694.). All participants gave written informed consent.

4.3.2 Exposure & clinical course measures

Several clinical outcomes were evaluated, including conversion to MS, occurrence of relapse and annualised disability progression from FDE to five-year review.

Conversion to MS was defined primarily as the occurrence of two or more clinical demyelinating episodes, thus satisfying the diagnostic requirements of dissemination in space and time, or a single episode plus paraclinical evidence, as per the 2005 McDonald criteria[12] (a minority of cases were diagnosed following MRI based on this latter criterion (n=20)). Conversion to MS was reported at annual review and cross-checked with neurological records. A relapse was defined according to the 2001 McDonald Criteria[13] as the acute or sub-acute appearance or reappearance of a neurological abnormality (lasting at least 24 hours) in the absence of other potential explanatory factors. Relapses were reported at annual review and only relapses which were diagnosed and verified by a neurologist were included in the analysis. Disability was assessed by the Kurtzke Expanded Disability Status Scale (EDSS)[14], assessed at the 5-year review; the EDSS on the day before FDE was assumed to be 0.

Clinical history was derived from medical records at initial presentation, describing the nature of the episode/symptoms which brought the participant into the Ausimmune Study, as well as historical symptoms prior to presentation. In the event that a person had no history preceding their referral symptoms, the referral symptom onset date was taken to be their symptom onset. Where a person had a bout-onset presentation and had symptoms some time previous to the referral episode, this was validated to the extent possible from available clinical notes contemporaneous with the historical episode or taken as valid if judged to be so by the attending neurologist

at the referral episode. Finally, where a person was progressive from onset, symptom onset was defined as either the earliest onset of symptoms identified by the attending neurologist at the referral episode, or one year preceding the referral clinic date, whichever was first.

Anti-EBNA-1 IgG, anti-EBNA-2 IgG and anti-HHV6 IgG titers were measured in serum samples collected at baseline, using immunofluorescence assays: Anti-EBNA1 commercial ELISA (DiaSorin), Anti-EBNA2 in house ELISA and Anti-HHV6 commercial ELISA (Panbio) as previously described[15].

4.3.3 Genotyping

DNA from AusLong participants was genotyped using the Illumina customised MS exome genotyping array (Illumina Human Exome-12 v1.2 array (~244,000 SNPs) plus additional MS relevant variants (~87,000) added as a customized component). Genotypes were called using Illumina GenomeStudio software. Individuals were excluded for the following reasons: a call rate of <99% or duplicate discordance. Variants were excluded on the basis of a call rate of <99% or a deviation from Hardy-Weinberg equilibrium with $P<1.0x10^{-6}$. Principal components analysis was carried out twice, once excluding HapMap samples to identify population outliers and once including HapMap samples to help interpret outliers[16]. Eight tagSNPs (rs9676113, rs3794832, rs7232502, rs12959006, rs61742988, rs3900176, rs11150997 and rs7233242) with $r^2<0.1$ in the MBP gene (chr18: 74690789-74844774) were selected for analysis.

4.3.4 Data analysis

Predictors of time to conversion to MS and of relapse were evaluated by Cox proportional hazards regression models, the latter for repeated events using the gap-time model by Prentice and colleagues[17]. All covariates satisfied the proportional hazards assumption.

Annualised change in EDSS was calculated by taking the five-year review EDSS and dividing by the duration between the day before FDE and the five-year review. Predictors of annualised change in EDSS were evaluated using linear regression, adjusted for whether persons were having a relapse at the time of their 5-year EDSS assessment. There were only 22 in total 279 samples that have active relapse at 5-year review. The average EDSS at 5-year review for all the samples vs no active relapse was 1.87 vs 1.82, such littler difference would not skew the results, thus adjusting for relapse at the time of 5-year assessment is adequate. Because the annualised change in disability was highly skewed, a log-transformation was applied to satisfy linear regression assumptions of minimal heteroskedasticity. All means and coefficients, however, were back-transformed and presented on the original scale of the change in EDSS.

Interaction was assessed by generating a product term of the two variables to be assessed, with the p-value of this two-component term delineating the significance of the interaction. All statistical analyses were conducted in Stata/SE 12.1 (StataCorp LP, College Station, Texas, USA).

4.4 Results

Of the 169 case participants in the Ausimmune/AusLong Study that had a classic FDE, 127 have undergone genotyping as described and had been assessed at 5 years and form the cohort assessed in this study. They were predominantly female (n=98, 77.2%), age at study entry (mean: 37.8, SD: 9.5), conversion to MS (n=68, 53.5%), relapse number (n=152) and 5-year EDSS (median: 1, IQR: 0 – 2) (Table 4.1).

Table 4.1: Characteristics of participants with genotype data

Characteristics	Classic FDE (n=127)
Female (%)	98(77.2)
Conversion to MS (%) [#]	68(53.5)
Relapse	152
Age at study entry, mean (SD)	37.8 (9.5)
EDSS, Median (IQR)	1 (0-2)

FDE: First demyelinating event

SD: Standard deviation, EDSS: Expanded Disability Status Scale

IQR: Interquartile range

[#] Participants who had CDMS and with genotype data.

4.4.1 Risk genotype (CT+TT) of rs12959006 directly predicts progression to MS, relapse and annualised change in EDSS

Among the group of participants with a classic FDE and a second event diagnostic of MS within the period after referral to the study, only two persons had the TT genotype of rs12959006. Therefore, the homozygous risk genotype was combined with the heterozygote as CT+TT. We found the genotype (CT+TT) of rs12959006 showed a borderline positive association with risk of conversion to MS (HR=1.57

(95% CI: 0.93-2.64), p=0.09). Examining this SNP in predicting relapse, a stronger and statistically significant result was seen (HR=1.74 (95% CI: 1.19-2.56), p=0.005, Table 4.2), partly reflecting the greater number of events (n=68 CDMS vs 152 relapses). Figure 4.2 showed the survival curves for the rs12959006 genotype for time to conversion to MS and relapse, respectively.

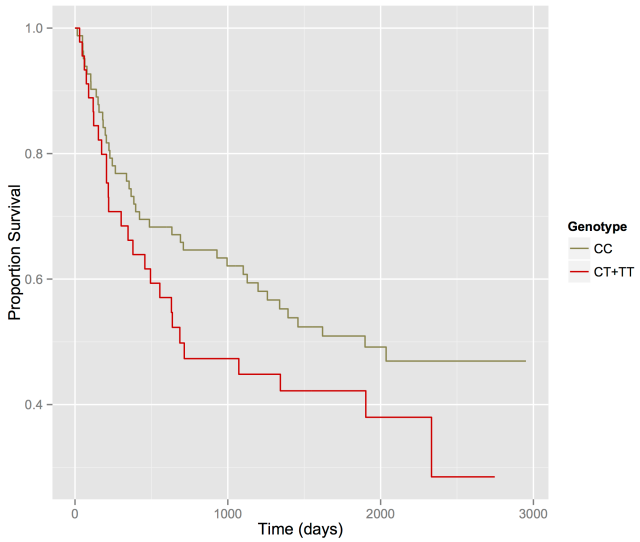
Table 4.2: Results for rs12959006 predicting MS, relapse and annualised ΔEDSS in MS among participants with a classic FDE.

rs12959006	N	Conversion to MS	N	Relapse	N	ΔEDSS
CC	41	Ref	70	Ref	37	0.26 (0.20, 0.32)
CT+TT	27	1.57 (0.93, 2.64)	82	**1.74 (1.19, 2.56)**	24	**+0.18 (0.06, 0.30)**
		p=0.09		**p=0.005**		**p=0.004**

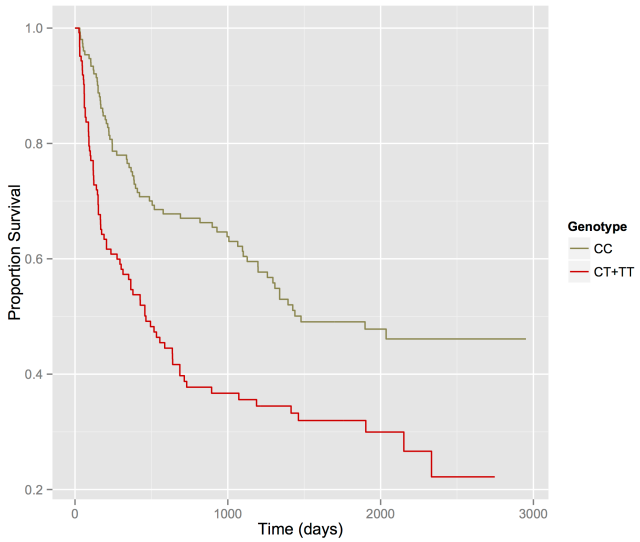
Due to the smaller number of people carrying TT genotype who converted to MS (N=2), we recoded the genotype as CT+TT. Results were adjusted for age, sex and study site, and presented as HR (95% CI) for time to MS and relapse. Disability results are presented as geometric mean annualised disability progression (95% CI) for the reference group, while coefficient relative to reference (β (95% CI)) are presented for subsequent levels. N refers to the number of events for each related clinical course.

Figure 4.2:

a. Kaplan-Meier survival plot for time to MS by category of rs12959006 genotype.



b. Kaplan-Meier survival plot for time to relapse by category of rs12959006 genotype.



Evaluating the association of rs12959006 with annualised change in EDSS, the CT+TT genotype was associated with a significantly greater rate of disability progression, with 0.18 greater annualised EDSS increase per year (β=0.18, 95% CI: 0.06-0.30, p=0.004). Further when assessing all cases with bout-onset disease (n=125), we observed a dose-dependent association for rs12959006 in predicting annualised change in EDSS, in which the CT genotype increased EDSS by 0.07 unit per year, while the TT genotype increased EDSS by 0.25 unit per year, both relative to the reference genotype (p_{trend}=0.02, Table 4.3). Translating this to a clinical outcome, those carrying the TT genotype will have an EDSS score 2.5 points greater over 10 years than those carrying the CC genotype.

Table 4.3: Results for SNP rs12959006 predicting annualised ΔEDSS in MS including all bout-onset cases.

rs12959006	N	ΔEDSS
CC	74	0.26 (0.21, 0.31)
CT	46	+0.07 (-0.02, 0.15)
TT	5	+0.25 (-0.02, 0.53)
		p_{trend}=0.02

Results were adjusted for age, sex and study site. Disability results are presented as the geometric mean annualised change in EDSS (95% CI) for the reference group; the β (95% CI) coefficient represents the difference relative to the reference for other genotypes. N refers to the number of events

We did not observe any association between the other seven MBP SNPs and progression to MS, relapse or annualised change in EDSS (data not shown). After adjusting for multiple testing ((p=0.05/8)≈0.006 was defined as significant), rs12959006 remained significant in predicting relapse and annualised ΔEDSS. The effect of the risk allele rs12959006 was in the same direction when predicting all

three clinical outcomes and was towards a more active clinical course, providing further support for a true effect.

4.4.2 Rs12959006 genotype interacts significantly with baseline anti-HHV6 IgG levels to predict conversion to MS and relapse

Table 4.4 : Results for SNP rs12959006 interact with baseline HHV-6, baseline EBNA-1 and baseline EBNA-2 to predict conversion to MS and relapse in MS among classic FDEs.

Factor	Genotype	CDMS			Relapse		
		N	HR (95% CI)	P	N	HR (95% CI)	P
Baseline HHV6	CC	23	0.63 (0.14, 2.77)	0.54	27	0.58 (0.20, 1.64)	0.3
	CT+TT	16	**6.95 (1.11, 43.31)**	**0.04**	**23**	**3.00 (1.19, 7.53)**	**0.02**
			$p_{interaction}$	**0.05**		**$p_{interaction}$**	**0.02**
Baseline EBNA1	CC	23	1.30 (0.57, 2.99)	0.54	27	1.10 (0.57, 2.12)	0.77
	CT+TT	16	**3.01 (1.15, 7.87)**	**0.02**	23	1.28 (0.88, 1.87)	0.19
			$p_{interaction}$	0.28		$p_{interaction}$	0.66
Baseline EBNA2	CC	23	1.12 (0.59, 2.14)	0.73	27	1.13 (0.74, 1.73)	0.57
	CT+TT	16	**4.14 (1.62, 10.56)**	**0.003**	23	1.52(0.90, 2.56)	0.12
			$p_{interaction}$	0.19		$p_{interaction}$	0.68

Due to the smaller number people carrying TT genotype converted to CDMS (N=2), we recoded the genotype as CT+TT. Results were adjusted for age, sex and study site, and presented as HR (95% CI) for conversion to MS and relapse. N refers to the number of events for each related clinical course.

As shown in Table 4.4, we found a significant interaction between the rs12959006 genotype and baseline anti-HHV6 IgG levels. In those with the risk genotype (CT+TT), there was a significant positive association between anti-HHV6 IgG and time to MS (HR=6.95, 95% CI: 1.11-43.31, p=0.04; $p_{interaction}$=0.05) and relapse (HR=3.00, 95% CI: 1.19-7.53, p=0.02; $p_{interaction}$=0.02), whereas there was no association amongst the non-risk allele carriers (CC) (HR=0.63, 95% CI: 0.14-2.77, p=0.54, for time to MS and HR=0.58, 95% CI: 0.20-1.64, p=0.30 for time to relapse). There was no significant interaction between rs12959006 genotype and baseline anti-

ENBA-1 or anti-EBNA-2 IgG titres for time to MS or relapse, although the effect was in the same direction as for HHV6. Rs12959006 genotype did not significantly interact with baseline anti-HHV6, anti-EBNA1 or anti-EBNA2 IgG levels in predicting annualised ΔEDSS.

Functional prediction analysis[18] using SNPinfo web server (https://snpinfo.niehs.nih.gov) showed this variant rs12959006 is the target of many transcription factors and the binding sites of miR-218 and miR-188-3p.

4.4.3 Discussion

Multiple sclerosis causes a great economic burden and has significant social, psychological and physical impacts on people with MS as well as their families. Following an FDE there are few good indicators of future conversion to MS or subsequent rate of progression, although early treatment may significantly alter outcomes. Treatment is expensive and for those with a good prognosis and low risk of conversion to MS almost certainly not warranted. Therefore identifying those with a higher risk of conversion and a faster rate of progression as early as possible in the disease course is of great importance[2,13,19]. The importance of MRI in early MS has been demonstrated[19]. However, other biomarkers have shown less promise, although serum antibodies against MBP have been associated with a significantly higher risk of relapse post-FDE[2].

We have shown that people with an FDE who carry the risk SNP of rs12969006 in *MBP* have a worse outcome on all three measures of disease progression (conversion to MS, relapse rate, and EDSS progression), and that the risk SNP of rs129569006 in

MBP interacts with serological markers of prior HHV6 infection to predict clinical course post FDE. Past infection with HHV6 and EBV are both well-recognised risk factors for MS onset and there is some evidence that HHV6 IgG levels in particular are associated with MS progression[20].

MBP undergoes complex post-transcriptional modification, including methylation, phosphorylation and miRNA binding[21]. The risk locus studied is a target for many transcription factors as well as the binding sites of miR-218 and miR-188-3p. Other research has shown that miR-218 expression is significantly down regulated in MS white matter compared to controls[22]. However, the exact molecular mechanisms by which changes in miR-218 and this MBP variant may modify myelination and demyelination remains unknown. Further studies using the deep sequencing to fine-mapping the *MBP* causal variants are needed.

Importantly we have shown that in the post-GWAS MS world, where the focus has shifted from defining risk associations to defining determinants of clinical course, studying *a priori* hypotheses such as those we have described, can be undertaken successfully in moderate sized, well-characterised longitudinal cohorts, using data on multiple aspects of MS clinical course and potential genetic and environmental factors.

There are several potential caveats to our findings. By their nature, longitudinal cohort studies are at best of moderate sample size, and further subdivision into those with particular phenotypes and exposure parameters decreases the power to detect associations. On the other hand, the internal consistency across the three outcome measures, the biologically plausible directions of effect and interactions with

herpesvirus, as well as the dose-dependency of effect in our disability analysis, is evidence supporting a true association rather than statistical artifact. Still, validation of our findings in other longitudinal cohorts is essential.

These data demonstrate that genetic variants in MS candidate genes with high biologic plausibility may help predict MS clinical course. Additionally, interaction with well-documented serological markers of prior EBV and HHV6 infection enhance these effects and further supports the notion of complex gene-environment interactions in the onset and progression of MS. These results, if replicated in other longitudinal datasets, may aid in developing prognostic algorithms in the early disease period in MS as well as providing further mechanistic insights.

4.5 Summary

Background: Prognosis following a first demyelinating event is difficult to predict, with no genetic markers of MS progression currently identified. MBP is a major component of the myelin sheath of CNS neurons and may play a central role in demyelinating diseases such as MS. However, genetic variation in MBP has not been implicated in MS onset risk in large genome-wide association studies. We hypothesised that genetic variations in MBP may be a determinant of MS clinical course.

Methods: We investigated whether variations in the MBP gene altered clinical course (conversion to MS and/or relapse, and annualised change in disability), using a longitudinal cohort study of persons who had had a first demyelinating event, followed up to the 5-year review.

Results: We found one variant, rs12959006, predicted worse clinical outcomes. The risk genotype (CT+TT) was near-significantly associated with increased risk of conversion to MS (HR=1.57, 95% CI=0.93-2.64, p=0.09); and significantly associated with relapse (HR=1.74, 95% CI=1.19-2.56, p=0.005) and of greater annualised disability progression (β=0.18, 95% CI=0.06-0.30, p=0.004). We also found a significant interaction between the risk genotype and baseline anti-HHV6 IgG in predicting MS (p interaction=0.05) and relapse (p interaction=0.02). Functional prediction analysis showed this variant is the target of many transcription factors and the binding sites of miR-218 and miR-188-3p.

Interpretation: Our results provide novel insights into the role of genetic variation within the MBP gene predicting MS clinical course, both directly and by interaction with known environmental MS risk factors.

4.6 References

1. Deber CM, Reynolds SJ. Central nervous system myelin: structure, function, and pathology. *Clin. Biochem.* 1991;24(2):113-134.

2. Berger T, Rubner P, Schautzer F, et al. Antimyelin antibodies as a predictor of clinically definite multiple sclerosis after a first demyelinating event. *N. Engl. J. Med.* 2003;349(2):139-145.

3. International Multiple Sclerosis Genetics C, Beecham AH, Patsopoulos NA, et al. Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. *Nat. Genet.* 2013;45(11):1353-1360.

4. International Multiple Sclerosis Genetics C, Wellcome Trust Case Control C, Sawcer S, et al. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature.* 2011;476(7359):214-219.

5. Stinissen P, Hellings N. Activation of myelin reactive T cells in multiple sclerosis: a possible role for T cell degeneracy? *Eur. J. Immunol.* 2008;38(5):1190-1193.

6. Holmoy T, Vartdal F. Infectious causes of multiple sclerosis. *Lancet Neurol.* 2005;4(5):268; discussion 269.

7. Tejada-Simon MV, Zang YC, Hong J, Rivera VM, Zhang JZ. Cross-reactivity with myelin basic protein and human herpesvirus-6 in multiple sclerosis. *Ann. Neurol.* 2003;53(2):189-197.

8. Soldan SS, Berti R, Salem N, et al. Association of human herpes virus 6 (HHV-6) with multiple sclerosis: increased IgM response to HHV-6 early antigen and detection of serum HHV-6 DNA. *Nat. Med.* 1997;3(12):1394-1397.

9. Pender MP, Burrows SR. Epstein-Barr virus and multiple sclerosis: potential opportunities for immunotherapy. *Clin Transl Immunology.* 2014;3(10):e27.

10. Pette M, Fujita K, Kitze B, et al. Myelin basic protein-specific T lymphocyte lines from MS patients and healthy individuals. *Neurology.* 1990;40(11):1770-1776.

11. Lucas R, Ponsonby AL, McMichael A, et al. Observational analytic studies in multiple sclerosis: controlling bias through study design and conduct. The Australian Multicentre Study of Environment and Immune Function. *Mult. Scler.* 2007;13(7):827-839.

12. Polman CH, Reingold SC, Banwell B, et al. Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. *Ann. Neurol.* 2011;69(2):292-302.

13. McDonald WI, Compston A, Edan G, et al. Recommended diagnostic criteria for multiple sclerosis: guidelines from the International Panel on the diagnosis of multiple sclerosis. *Ann Neurol.* 2001;50(1):121-127.

14. Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). *Neurology.* 1983;33(11):1444-1452.

15. Strautins K, Tschochner M, James I, et al. Combining HLA-DR risk alleles and anti-Epstein-Barr virus antibody profiles to stratify multiple sclerosis risk. *Mult. Scler.* 2014;20(3):286-294.

16. Guo Y, He J, Zhao S, et al. Illumina human exome genotyping array clustering and quality control. *Nat. Protoc.* 2014;9(11):2643-2662.

17. Prentice RL, Williams BJ, Peterson AV. On the regression analysis of multivariate failure time data. *Biometrika.* 1981;68(2):373-379.

18. Xu Z, Taylor JA. SNPinfo: integrating GWAS and candidate gene information into functional SNP selection for genetic association studies. *Nucleic Acids Res.* 2009;37(Web Server issue):W600-605.

19. O'Riordan JI, Thompson AJ, Kingsley DP, et al. The prognostic value of brain MRI in clinically isolated syndromes of the CNS. A 10-year follow-up. *Brain.* 1998;121 (Pt 3):495-503.

20. Simpson S, Jr., Taylor B, Dwyer DE, et al. Anti-HHV-6 IgG titer significantly predicts subsequent relapse risk in multiple sclerosis. *Mult. Scler.* 2012;18(6):799-806.

21. Harauz G, Ishiyama N, Hill CM, Bates IR, Libich DS, Fares C. Myelin basic protein-diverse conformational states of an intrinsically unstructured protein and its roles in myelin assembly and multiple sclerosis. *Micron.* 2004;35(7):503-542.

22. Noorbakhsh F, Ellestad KK, Maingat F, et al. Impaired neurosteroid synthesis in multiple sclerosis. *Brain.* 2011;134(Pt 9):2703-2721.

Supplementary Table 4-1: Results for other 7 MBP SNPs predicting CDMS, relapse and annualised ΔEDSS in MS among the classic FDEs.

SNP	Genoypte	N	CDMS(HR(95%CI))	N	Relapse(HR(95%CI))	N	ΔEDSS (β (95%CI))
rs9676113	AA	39	Ref	80	Ref	36	Ref
	AG+GG	29	0.72(0.43,1.20)	71	1.10(0.70,1.69)	25	0.03(-0.09,0.14)
			p=0.20		p=0.70		p=0.65
rs3794832	GG	65	NA	145	Ref	59	NA
	AG	3	NA	6	1.06(0.21,5.27)	2	NA
			NA		p=0.94		NA
rs7232502	GG	28	Ref	64	Ref	24	Ref
	AG	32	0.94(0.55,1.60)	73	0.85(0.54,1.32)	30	-0.08(-0.21,0.05)
	AA	8	1.16(0.45,2.98)	14	0.84(0.43,1.64)	7	-0.13(-0.31,0.04)
			p=0.90		p=0.48		p=0.12
rs61742988	GG	66	NA	144	Ref	59	NA
	AG	2	NA	7	0.92(0.33,2.55)	2	NA
			NA		p=0.88		NA
rs3900176	GG	27	Ref	62	Ref	25	Ref
	GA	31	1.12(0.66,1.92)	68	0.91(0.56,1.46)	26	-0.01(-0.14,0.12)
	AA	10	1.57(0.75,3.31)	22	0.97(0.52,1.78)	10	-0.06(-0.23,0.10)
			p=0.28		p=0.81		p=0.49
rs11150997	GG	67	NA	149	NA	60	NA
	GA	1	NA	2	NA	1	NA
			NA		NA		NA

rs7233242	GG	26	Ref	62	Ref	24	Ref
	GA	31	0.97(0.56,1.68)	79	1.14(0.73,1.79)	27	0.05(-0.08,0.17)
	AA	11	1.11(0.51,2.40)	11	0.70(0.38,1.30)	10	-0.12(-0.27,0.02)
			p=0.86		p=0.73		p=0.30

For cell with number <5, the analysis was not performed and annotated with NA. Results were adjusted for age, gender and studysite, and presented as HR (95% CI) for CDMS and relapse, and β (95%CI) for EDSS. N refers to the number of events for each related clinical course.

Chapter 5: Common genetic variation within *miR-146a* predicts disease onset & relapse in multiple sclerosis

5.1 Preface

This chapter will address the key question discussed in the Chapter 1 section regarding the role of epigenetics in the missing heritability of MS. In Chapter 5, for the first time, we assess the effects of a functional genetic variant within the gene encoding *miR-146a* (a miRNA with expression strongly associated with MS disease activity) on the MS clinical course in the Ausimmune/Auslong Longitudinal Study—an ongoing > 10year prospective cohort study following a population of 282 patients who were initially identified with a first demyelinating event (FDE) suggestive of, but not diagnostic for, MS. *miR-146a* expression is also known to interact with known MS risk factors, including Epstein Barr virus and smoking, making it a highly biologically plausible MS risk loci.

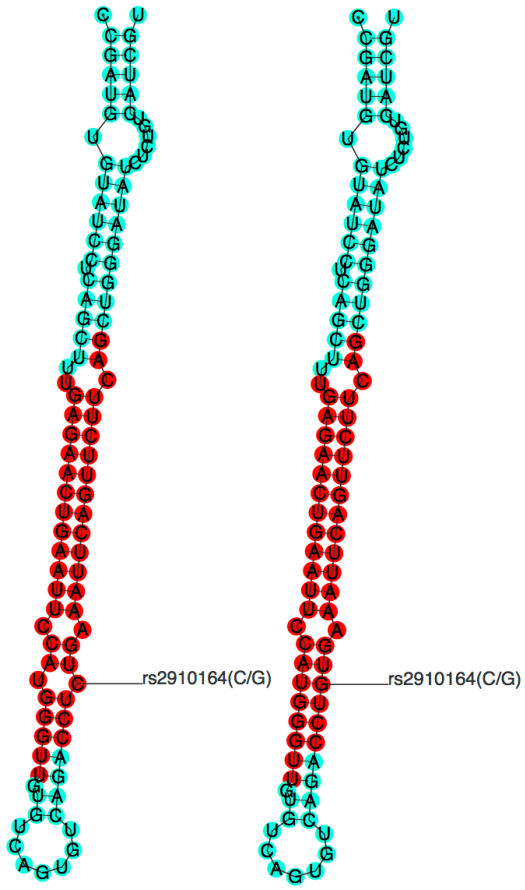
5.2 Introduction

MicroRNAs (miRNAs) are a family of 21 to 25 nucleotide-long noncoding small RNAs that regulate gene expression[1]. Despite extensive studies focusing on changes in the expression of miRNAs in multiple sclerosis (MS) compared to healthy controls[2], no studies have sought to determine the function of genetic variation in miRNAs genes, or to assess whether such variation could predict MS clinical course, directly or by interaction with other risk factors.

One particular miRNA (miR-146a) has been investigated for its role in MS due to its implicated in the regulation of both innate and adaptive immunity. MiR-146a is primarily involved in the regulation of inflammation[3] and it is up-regulated in active MS brain lesions[4]. In MS patients with relapsing remitting disease, *miR-146a* expression was significantly lower in those treated with glatiramer acetate than in treatment-naïve patients[5]. Additionally the expression of *miR-146a* is altered by infection with Epstein-Barr virus (EBV)[6] and smoking[7], both of which are well-known MS risk factors[8,9]. For these reasons, *miR-146a* is a biologically plausible candidate genetic risk locus in MS.

The minor allele C of rs2910164, which is the only common SNP located within the *miR-146a* gene (chr5: 159,912,359-159,912,457), causes mis-pairing within the hairpin of pre-miR-146a (Figure 5.1) and may decrease the expression of mature miR-146a[10] either by reducing the stability of the pri-miRNA, the efficiency of processing of pri-miRNA into pre-miRNA, or the efficiency of processing the pre-miRNA into the mature miRNA. Given the potential for both direct and interactive relationships of *miR-146a* with MS risk and clinical course, as well as the impact of this polymorphism on the expression and function of *miR-146a*, we developed an *a priori* hypothesis that this polymorphism may be an important modulator of MS clinical activity.

Figure 5.1: The predicted secondary structure of pre-miR-146a by rs2910164 polymorphism. The red represents the mature miR-146a duplex.



We therefore undertook a study of clinical outcomes in a well-established, prospectively followed cohort of patients who were first recruited after diagnosis of a first demyelinating event (FDE) or initial diagnosis of progressive disease, to assess the effect of this functional genetic variation in the *miR-146a* gene on the transition to MS, relapse frequency and disability progression.

5.3 Methods

5.3.1 Study design

Same as the study design section in Chapter 4.

5.3.2 Exposure & clinical course measures

Several clinical outcomes were evaluated as described in Chapter 4, including conversion to MS, occurrence of relapse, and annualized disability progression from FDE to five-year review.

5.3.3 Genotyping

Genotyping for rs2910164 was performed using the predesigned TaqMan® genotyping assay C_15946974_10 from ThermoFisher Scientific (catalog number 4351379). Samples were genotyped in 96-well plates on a Roche LightCycler® 480 Real-Time PCR system. Per sample 7 μL reaction volumes were used with 1 μL of DNA at 10 ng/μL, 4 μL of TaqMan® Genotyping Master Mix (Applied Biosystems™), 2.9 μL of nuclease free H_2O and 0.1 μL of 40× TaqMan genotyping probe. 5.9% of samples were genotyped in replicate with 100% replication.

5.3.4 Data analysis

The methods for predictors of time to conversion to MS, of relapse and of annualized change were the same as described in Chapter 4.

Interaction was assessed in two fashions, multiplicative and additive. Multiplicative interaction was assessed by generating a product term of the two variables to be assessed, with the significance of this two-component term delineating the interpretation of the significance of the interaction. Additive interaction was assessed by generating a four-level categorical term equal to 0 where both genetic and environmental risk factors were not present (the background risk (r_B)), equal to 1 and

2 where only the genetic or environmental risk factors were present (r_G and r_E), and equal to 3 where both genetic and environmental risk factors were present (r_{GE}). When evaluating a continuous environmental factor, a dichotomous variable was generated by dividing the sample at the median of the continuous term. The relative excess risk due to interaction (RERI) on the additive scale was calculated by subtracting the sum of risk measures (e.g. hazard ratios, HRs) for the individual exposures ($r_E + r_G$) from the risk measure for the combined exposure (r_{GE}) and adding the background risk (r_B =1.00) (i.e., RERI= $HR_{GE} – (HR_G + HR_E) + 1$). The statistical significance of this RERI (probability of a risk of this magnitude occurring due to chance alone) was evaluated by a permutation simulation, randomly redistributing participants to one of the four levels of the additive interaction term in proportion to their original distribution. For example, the proportions of the four-level EBNA1-miRNA interaction term (32.5%, 31.5%, 19.8%,16.2%) were retained in the simulation, but redistributed randomly amongst the participants who had data on both EBNA1 and miRNA genotype. These simulated interaction terms were generated, analysed and the magnitudes of the estimates resulting were retained. These simulations were run 50,000 times and the proportion of magnitudes resulting that were as or more extreme than that found in the as-measured analyses denoted the significance of the interaction.

All statistical analyses were conducted in Stata/SE 12.1 (StataCorp LP, College Station, Texas, USA).

5.4 Results

Of the 169 case participants in the Ausimmune/AusLong Study that had a classic FDE, 151 have undergone genotyping as described and had been assessed at 5 years and form the cohort assessed in this study. They were predominantly female (n=117, 77.5%), age at study entry (mean: 37.4, SD: 9.5), conversion to MS (n=83, 55.0%), relapse number (n=198) and 5-year EDSS (median: 1.25, IQR: 0 – 2) (Table 5.1).

Table 5.1: Characteristics of participants with genotype data

Characteristics	Classic FDE (n=151)
Female (%)	117(77.5)
Conversion to MS (%) [#]	83(55.0)
Relapse	198
Age at study entry, mean (SD)	37.4 (9.5)
EDSS, Median (IQR)	1.25 (0-2)

FDE: First demyelinating event

SD: Standard deviation, EDSS: Expanded Disability Status Scale

IQR: Interquartile range

[#] Participants who had CDMS and with genotype data.

5.4.1 Risk genotype (GC+CC) of rs2910164 directly predicts progression to relapse

Among the group of participants whose FDE occurred immediately prior to recruitment to the Ausimmune Study (n =151) and who were thus at risk of a second event diagnostic of MS within the period after referral to the study, we found that having genotype (GC+CC) of rs2910164 compared to GG was associated with an

increased hazard of converting to MS (HR: 1.52 (95% CI: 0.97, 2.42), p=0.07, Table 5.2), although not statistically significant.

Table 5.2: Association between miR-146a SNP rs2910164 genotype and measures of clinical course.

rs2910164	N	Conversion to MS (HR (95% CI))	n	Relapse (HR (95% CI))	n
GG	93	Ref	46	Ref	79
GC+CC	58	1.52 (0.97, 2.42)	37	**2.10(1.43,3.08)**	120
		p=0.071		**P=0.0001**	

Due to the small number of people carrying the CC genotype who converted to MS (N=1), we recoded the genotype as GC+CC. Results were adjusted for age, sex and study site, and presented as HR (95% CI) for time to MS and relapses. The N refers to the number of rs2910164 genotype in the subgroup with a "classic FDE". n refers to the number of related clinical course.

A stronger result was seen for relapse (HR=2.09 (95% CI: 1.42, 3.06), p=0.0001), partly reflecting the greater number of events (n=83 Conversion to MS vs. 198 relapses). Figure 5.2 and Figure 5.3 show the survival curves for the two genotypes for conversion to MS and relapse, respectively. The rs2910164 genotype did not predict annualised change in EDSS (p=0.25, Table 5.2).

Figure 5.2: Kaplan-Meier survival plot for time to conversion to MS by category of rs2910164 genotype for those with a classical FDE (n=151).

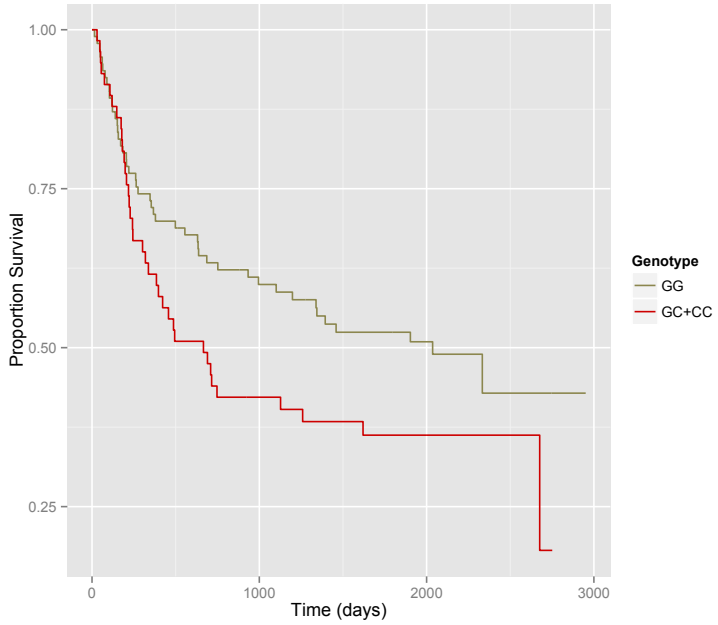
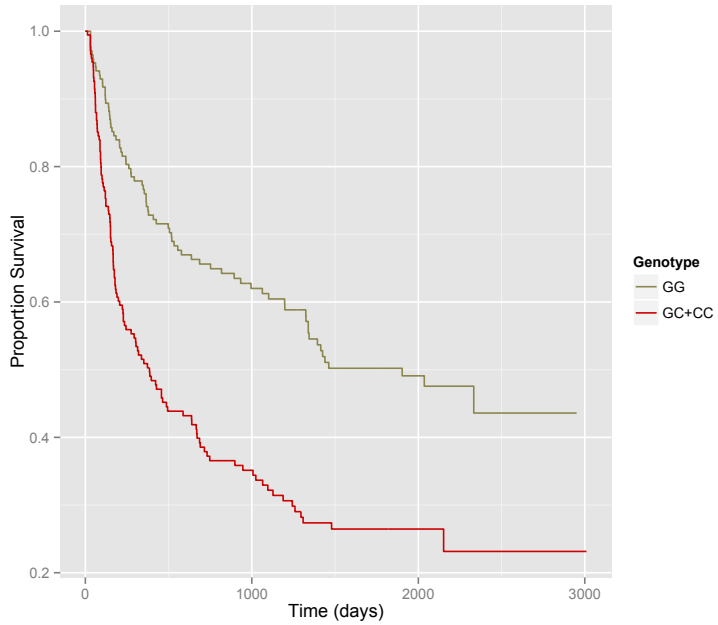


Figure 5.3: Kaplan-Meier survival plot for time to relapse by category of rs2910164 genotype.



5.4.2 Interactions of rs2910164 genotype with EBV related variables in predicting MS and relapse.

The rs2910164 SNP did not significantly interact with EBV-related parameters (serological, clinical) on the multiplicative scale in predicting either MS or relapse (Table 5.3). We did, however, find a significant interaction on the additive scale between rs2910164 and baseline anti-EBNA-1 IgG titers predicting risk of conversion to MS (RERI: 2.39, p=0.00002,Figure 5.4) and relapse (RERI: 1.20, p=0.006, Figure 5.5) (Table 5.3). The combined effect of having both a high baseline anti-EBNA-1-IgG titer and the risk genotype (GC+CC) of *miR-146a* was higher than expected based on the effects of having high baseline anti-EBNA-1-IgG titer in the absence of risk genotype (GC+CC) of *miR-146a*, and the effects of having risk genotype (GC+CC) of *miR-146a* in the absence of high baseline anti-EBNA-1-IgG titer. Supporting these results, similar results were seen for the other EBV correlated variables: anti-EBNA-2 IgG titers and having a history of infectious mononucleosis (Table 5.3). No significant interactions were observed with smoking (data not shown).

Figure 5.4: Hazard ratios for time to convert to MS by different combinations of baseline anti-EBNA-1 IgG titers (Low: <=0.53, High:>0.53) and miR-146a genotype (GG, GC+CC). Test for interaction on the multiplicative scale, HR=1.48,p=0.38; test for interaction on the additive scale, RERI=2.39, p=0.00002.

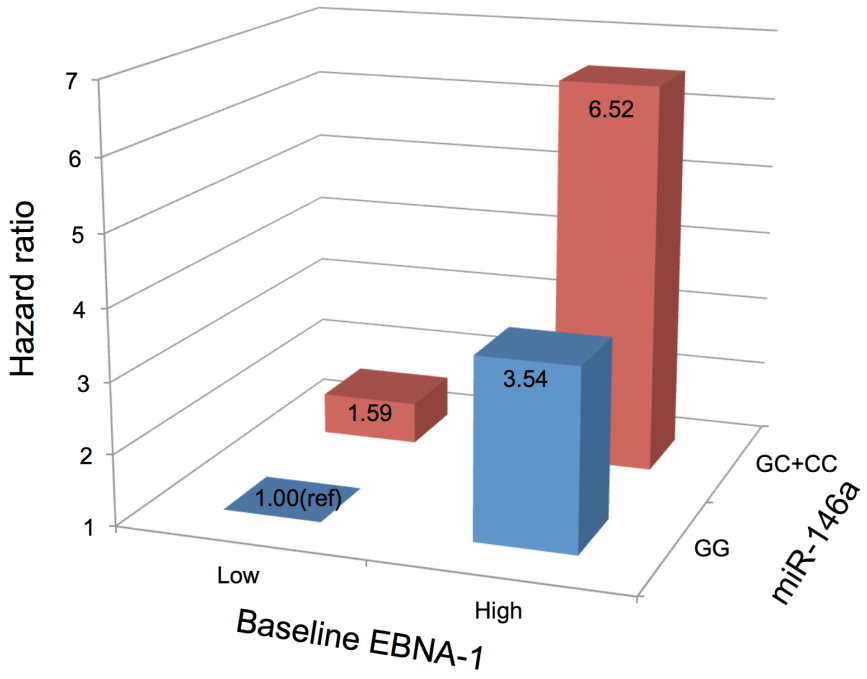


Figure 5.5: Hazard ratios for time relapse by different combinations of baseline anti-EBNA-1 IgG titers (Low: <=0.53, High:>0.53) and miR-146a genotype (GG, GC+CC). Test for interaction on the multiplicative scale, HR=1.16,p=0.82; test for interaction on the additive scale, RERI=1.20, p=0.006.

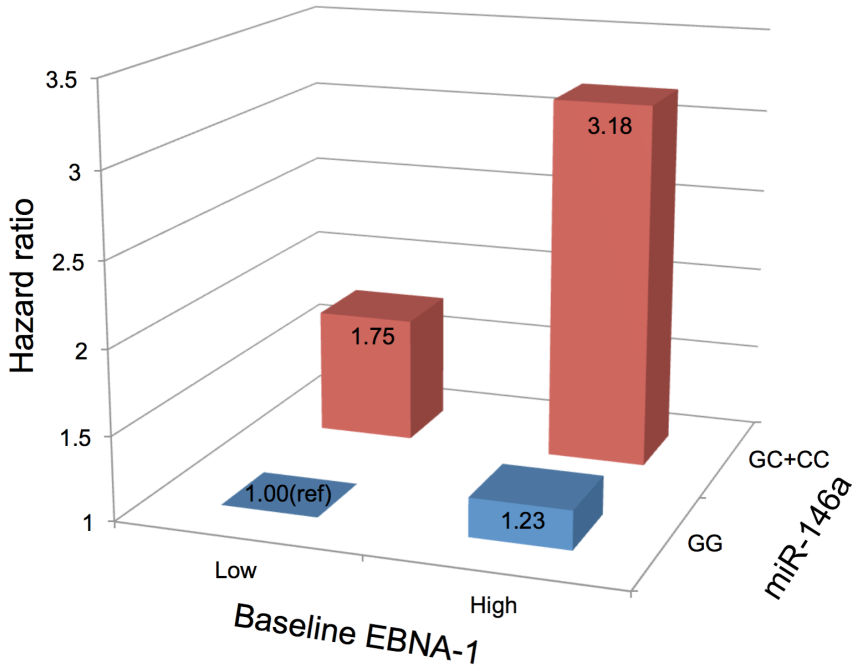


Table 5.3: Interaction between miR-146a SNP rs2910164 with baseline EBV antibody levels (anti-EBNA-1 and anti-EBNA-2) and according to history of having had infectious mononucleosis on time to conversion to MS and relapse.

Clinical Course	Factor E	Factor G	Effects of having factor E in absence of factor G	Effects of having factor G in absence of factor E	Effects of having both factor E and factor G	Interaction Parameters	
						Multiplicative Interaction	Additive Interaction
			HR_E(95%CI)	HR_G(95%CI)	HR_{GE}(95%CI)	HR_{multi}(p)	RERI(p)
Conversion to MS	High baseline EBNA1	miR-146a (GC+CC)	3.54(1.48-8.43)	1.59(0.54-4.67)	6.52(2.41-17.67)	1.48(0.38)	**2.39(0.00002)**
	High baseline EBNA2	miR-146a (GC+CC)	1.62(0.76,3.46)	1.09(0.40,2.94)	2.49(1.09,5.66)	1.41(0.60)	**0.78(0.05)**
	Infectious mononucleosis	miR-146a (GC+CC)	0.60(0.26,1.38)	1.35(0.79,2.31)	1.38(0.69,2.77)	1.69(0.34)	0.43(0.21)
Relapse	High baseline EBNA1	miR-146a (GC+CC)	1.23(0.69-2.17)	1.75(0.90-3.40)	3.18(1.69-5.97)	1.16(0.82)	**1.20(0.006)**
	High baseline EBNA2	miR-146a (GC+CC)	0.87(0.49,1.54)	1.22(0.68,2.19)	2.65(1.56,4.52)	2.50(0.03)	**1.56(0.005)**
	Infectious mononucleosis	miR-146a (GC+CC)	1.21(0.49,2.98)	1.72(1.15,2.58)	3.40(2.24,5.17)	2.53(0.09)	**1.47(0.0006)**

High baseline EBNA1 status (>0.53 vs <=0.53); High baseline EBNA2 status(>0.70 vs <=0.70); Infectious mononucleosis (yes vs no);miR146a (GC+CC) (GC+CC vs GG)

HR: hazard ratio; CI: confidence interval; RERI: relative excess risk due to interaction on additive scale

5.5 Discussion

Here we have shown that the risk genotype (GC+CC) of *miRNA-146a* SNP rs2910164 directly predicts conversion to MS and risk of relapse but not accumulation of disability. We also provide evidence of an interaction on the additive scale, between the risk genotype of rs2910164 and baseline-measured EBV-related serological and clinical variables in predicting MS and relapse, such that these environmental factors are significantly more strongly associated with clinical outcomes among those carrying the *miR-146a* risk variant.

The role of *miRNAs* in the development and intensity of inflammatory reactions in complex disorders such as MS is now being explored. Increased expression of *miR-146a* has been found to be associated with a more intense inflammatory state in MS brain lesions[4] and to be down-regulated by treatment with glatiramer acetate[5]. Jazdzewski et al found that the C allele of rs2910164 demonstrated lower expression compared with the non-risk G allele using transfected cells[11]. In MS, patients with the C allele of rs2910164 (compared to those with the GG and GC alleles) had higher expression of the prototypical pro-inflammatory cytokines TNF-α and IFN-γ in peripheral blood mononuclear cells[12]. Such changes were modulated by miR-146a through signaling pathways targeting *TRAF6* and *IRAK1* gene expression, which establishes a negative feedback loop that controls the intensity and the duration of NF-kB signaling. Recently, upregulated miR-146a in the brain of patients with Alzheimer's disease (AD) has been shown to significantly increase the severity of AD by directly binding the target site of LRP2 mRNA, resulting in decreased protein expression. LRP2 is expressed on the surface of many CNS cells including neurons and oligodendrocytes and is a critical receptor in axonal guidance. Therefore this

genetic variant that potentially alters expression of miR-146a is biologically plausible as a causal influence on the clinical course of MS.

Previous work has shown links between miR-146a and EBV infection. Using EBV de novo infection of primary cultured human B cells, miR-146a was down-regulated on initial infection, but was dramatically up-regulated upon induction of the lytic cycle[6]. In T-cells, decreased expression of miR-146 can cause hyper-responsiveness to T-cell receptor signals and unresolved T-cell-mediated inflammation[13]. Several studies have demonstrated the importance of T-cell driven immune responses in the association between EBV infection and onset of MS[14]. These effects of EBV infection on *miR-146a* expression, and on the effectiveness of T-cell responses, provide plausibility to enhanced effects on disease course of combined variation in *miR-146a* genotype and markers of EBV infection.

In our analyses we demonstrated an additive interaction between EBV infection and *miR-146a* genotype. Similar additive interactions have been previously demonstrated between markers of EBV infection and the *HLA-DRB1*15* genotype[15,16] and have been interpreted as indicating that the two factors are component causes of the same sufficient cause for a substantial proportion of people who are in the at-risk category of each factor – in this case the (GC+CC) genotype and a higher than median anti-EBNA-1 IgG level. We did not find a significant multiplicative interaction . We also did not observe interaction between smoking and *miR-146a* genotype. However, interaction between smoking and genetic factors in predicting MS risk has been controversial. In both of the above studies[15,16] there was no significant additive interaction between smoking and *HLA-DRB1*15* genotype in predicting MS risk. On

the contrary, work by Anna et al.[17] observed an additive interaction between passive smoking and carriage of *HLA-DRB1*15* risk genotype.

The strengths of our study include that we have detailed information on the clinical course from the first diagnosis of CNS demyelination for over five years, with high cohort retention. Only medically confirmed relapses were included in the analysis. One limitation is that the sample size of participants with a FDE close to the time of recruitment with no second event prior to data collection was relatively small, limiting the power of the study. A strength of our study is that in Australia treatment with disease modifying therapy (DMT) typically does not commence until a diagnosis of MS has been made. Our cohort is thus particularly valuable as the results on conversion to MS are not significantly confounded by treatment with DMT.

In conclusion, we provide evidence that a functional genetic variant in *miR-146a*, which has been linked *in vitro* with a change in *miR-146a* expression, predicts conversion to MS and relapse in MS. The combination of the risk genotype of *miR-146a* and higher levels of anti-EBNA-IgG result in a marked increase in risk of conversion to MS and relapse, suggesting that these two factors are component causes of a sufficient cause. Additional studies are needed to validate the findings and to further elucidate the pathways by which miR-146a, independently or in association with EBV infection, can influence disease progression following initial CNS demyelination.

5.6 Summary

Background: Despite extensive studies focusing on the changes in expression of miRNAs in MS compared to healthy controls, no studies have evaluated the association of genetic variants of miRNAs with MS clinical course.

Methods: We investigated whether a functional polymorphism in the MS associated *miR-146a* gene predicted clinical course (hazard of conversion to MS and/or relapse, and annualized change in disability), using a longitudinal cohort study of persons with a first demyelinating event followed up to their 5-year review.

Results: We found the genotype (GC+CC) of rs2910164 predicted relapse compared with the GG genotype (HR=2.09 (95% CI: 1.42, 3.06), p=0.0001), as well as a near-significant (p=0.07) association with conversion to MS. Moreover, we found a significant additive interaction between rs2910164 and baseline anti-EBNA-1 IgG titers predicting risk of conversion to MS (RERI: 2.39, p=0.00002) and of relapse (RERI: 1.20, p=0.006). Supporting these results, similar results were seen for the other EBV-correlated variables: anti-EBNA-2 IgG titers and past history of infectious mononucleosis. There was no association of rs2910164 genotype for disability progression.

Conclusion: Our findings suggest *miR-146a* and its target genes are potential therapeutic targets for modulating clinical course in MS.

5.7 References:

1. Baek D, Villen J, Shin C, Camargo FD, Gygi SP, Bartel DP. The impact of microRNAs on protein output. *Nature.* 2008;455(7209):64-71.

2. Zhou Y, Simpson S, Jr., Holloway AF, Charlesworth J, van der Mei I, Taylor BV. The potential role of epigenetic modifications in the heritability of multiple sclerosis. *Mult Scler.* 2014;20(2):135-140.

3. Liu G, Abraham E. MicroRNAs in immune response and macrophage polarization. *Arterioscler. Thromb. Vasc. Biol.* 2013;33(2):170-177.

4. Junker A, Krumbholz M, Eisele S, et al. MicroRNA profiling of multiple sclerosis lesions identifies modulators of the regulatory protein CD47. *Brain.* 2009;132(Pt 12):3342-3352.

5. Waschbisch A, Atiya M, Linker RA, Potapov S, Schwab S, Derfuss T. Glatiramer acetate treatment normalizes deregulated microRNA expression in relapsing remitting multiple sclerosis. *PLoS One.* 2011;6(9):e24604.

6. Godshalk SE, Bhaduri-McIntosh S, Slack FJ. Epstein-Barr virus-mediated dysregulation of human microRNA expression. *Cell Cycle.* 2008;7(22):3595-3600.

7. Vrijens K, Bollati V, Nawrot TS. MicroRNAs as Potential Signatures of Environmental Exposure or Effect: A Systematic Review. *Environ. Health Perspect.* 2015.

8. Ascherio A, Munger KL. Environmental risk factors for multiple sclerosis. Part I: the role of infection. *Ann. Neurol.* 2007;61(4):288-299.

9. Ascherio A, Munger KL. Environmental risk factors for multiple sclerosis. Part II: Noninfectious factors. *Ann. Neurol.* 2007;61(6):504-513.

10. Gong J, Tong Y, Zhang HM, et al. Genome-wide identification of SNPs in microRNA genes and the SNP effects on microRNA target binding and biogenesis. *Hum. Mutat.* 2012;33(1):254-263.

11. Jazdzewski K, Murray EL, Franssila K, Jarzab B, Schoenberg DR, de la Chapelle A. Common SNP in pre-miR-146a decreases mature miR expression and predisposes to papillary thyroid carcinoma. *Proc. Natl. Acad. Sci. U. S. A.* 2008;105(20):7269-7274.

12. Li Y, Du C, Wang W, et al. Genetic association of MiR-146a with multiple sclerosis susceptibility in the Chinese population. *Cell. Physiol. Biochem.* 2015;35(1):281-291.

13. Baumjohann D, Ansel KM. MicroRNA-mediated regulation of T helper cell differentiation and plasticity. *Nat. Rev. Immunol.* 2013;13(9):666-678.

14. Pender MP, Burrows SR. Epstein-Barr virus and multiple sclerosis: potential opportunities for immunotherapy. *Clin Transl Immunology.* 2014;3(10):e27.

15. Disanto G, Hall C, Lucas R, et al. Assessing interactions between HLA-DRB1*15 and infectious mononucleosis on the risk of multiple sclerosis. *Mult. Scler.* 2013;19(10):1355-1358.

16. Sundqvist E, Sundstrom P, Linden M, et al. Lack of replication of interaction between EBNA1 IgG and smoking in risk for multiple sclerosis. *Neurology.* 2012;79(13):1363-1368.

17. Hedstrom AK, Bomfim IL, Barcellos LF, et al. Interaction between passive smoking and two HLA genes with regard to multiple sclerosis risk. *Int. J. Epidemiol.* 2014;43(6):1791-1798.

Chapter 6: Conclusion

Multiple sclerosis is a complex disease of the central nervous system in which both environmental and genetic factors contribute to disease development. The incidence of MS is increasing every year, creating a huge economic burden for the patients and their families. Despite the increasing understanding of MS, no cure exists. Work is ongoing to complete the puzzle that is the etiology of MS, and the ultimate goal of research is to identify a cure for the disease. The work presented in this thesis provides further insights into the missing heritability in MS using a variety of study designs and methodologies. This chapter will summarize some key conclusions from the work and provide directions for further research.

6.1 Genetic loci associated with EBV infection and their roles in MS risk

Infection with EBV has been implicated in the development of MS. Despite extensive studies demonstrating the association between anti-EBNA-1 IgG titers and MS risk, few studies have investigated the susceptibility genes associated with anti-EBNA-1 IgG titers and their roles in MS risk. In Chapter 2, we performed a GWAS using QIMR twin families by measuring the anti-EBNA-1 IgG titers in the samples and treating them as a quantitative trait. Twin studies offer a powerful approach to the identification of genetic variation underlying complex traits due to the shared genetic and environmental exposures between twins[1]. With twin family data available by collaborating with other research groups and given the high rate of EBV infection in the general population[2] (>90%), GWAS becomes relatively cheaper, as opposed to

collecting a larger number of MS samples, measuring the anti-EBNA-1 IgG titers, and conducting a GWAS.

Using this approach, we identified a strong association with the HLA region for the significant SNP, rs2516049, located between genes *HLA-DRB1 and HLA-DQA1* in the QIMR twin families. In the conditional analysis, adjusting for the top significant SNP as a covariate did not demonstrate any other independent signals located within the HLA regions, suggesting that these SNPs are within one haplotype block containing the variants modulating anti-EBNA-1 IgG titers. Additionally, we did not observe any significant loci outside the HLA region. In an effort to identify novel non-HLA region loci, we then performed a meta-analysis using data from another EBNA-1 GWAS from a cohort of Mexican-American families. Again, we only observed a significant signal within the HLA region.

From this analysis, the question remains: how do genetic loci associated with anti-EBNA-1 IgG titers modulate MS risk? To answer this question, we examined the shared polygenetic risk between the EBNA-1 GWAS and the largest published MS GWAS by applying a novel method called SNP effect concordance analysis (SECA)[3], and we then replicated the SECA findings using the genetic risk score method. We found shared genetic risk between the EBNA-1 GWAS and MS GWAS. In the meta-analysis combining the two datasets, we found that non-HLA loci at 1p22.1, 3p24.1, 3q13.33, and 10p15.1 reached genome-wide significance ($P < 5\times10^{-8}$). In line with functional research work, our results support a role for these loci in EBV infection and thus in contributing to MS development. As an example, the gene *Eomesodermin* (*EOMES*), located on 3p24.1, has been reported in a study performed by Aalderen and colleagues[4]. The authors found direct evidence that *EOMES* expression was altered

after EBV infection. Similarly, Parnell and colleagues[5] showed that *EOMES* expression was significantly lower in MS cases compared with healthy controls. Thus, our results here provide guidance for further study aimed at analyzing these loci, which may yield a better understanding of the possible molecular mechanisms involved in these processes.

Furthermore, some of the study designs and methods that we described in this chapter can be further investigated in future research. First, the twin study design can be applied to studies analyzing other environmental/behavioral variables, and then the genes discovered from this stage can be used as a priori hypotheses for testing in the diseases that are caused by such environmental/behavioral variables. Second, shared polygenetic risk analysis can be applied to examine the shared risk between different diseases. For example, it is known that Alzheimer's[6], Parkinson's[7], and MS are all neurodegenerative diseases. However, despite extensive GWAS studies looking for susceptibility loci for each disease separately, few studies have sought to ascertain whether there is shared genetic risk among these diseases[8-12]. Our method may provide a better approach, potentially leading to the discovery of shared genetic determinants involved in the pathways contributing to these diseases.

6.2 Genetic variations within cytokines and their receptor genes in predicting MS

Abnormal cytokine production by peripheral blood mononuclear cells (PBMC) plays an important role in the pathogenesis of MS[13]. In previous work, we showed that increasing levels of IFN-γ were significantly associated with increased risk of relapse, whereas TNF-α showed a significant inverse relationship with relapse risk[14]. In

Chapter 3, we further expanded this analysis to systematically assess whether the variations within the network of cytokine genes and their receptor genes modulate the association of IFN-γ and TNF-α with relapse. All of the known cytokine genes and their receptors genes (N=83) within a gene boundary of 10 kb containing 361 SNPs were selected for analysis. We showed that rs3218295 (within the gene *IL2RB*) and rs522807 (3' of the gene *TNFRSF1B*) showed a significant interaction with TNF-α as a predictor of relapse, whereas rs25879 (5' of the gene *IL3*) and rs522807 (3' of the gene *TNFRSF1B*) showed a significant interaction between IFN-γ and relapse.

Our results demonstrated potential roles for *IL2RB* and *IL3* in the development of MS and provided insight into the complex interactions between genetic variations and the cytokines in predicting clinical course. Additional evidence has supported a plausible biological role for these genes in MS. For example, SNP rs31480, which has high LD with our significant SNP, rs25879, significantly influenced IL3 gene expression, which then affects the proliferation and survival of neural progenitors[15]. Although there have been no studies showing a role for *IL2RB* in MS[16], this gene has been identified as a susceptibility locus for other autoimmune disease such as rheumatoid arthritis[17]. Thus, additional studies are needed to better understand its role in MS.

Importantly, the genes identified as interacting with PBMC-produced cytokines in predicting clinical course were not identified as susceptibility loci in previous GWAS studies. This may be because GWAS typically use a cross-sectional study design and are limited by multiple testing burdens. Therefore, GWAS is insensitive in identifying genetic predictors in clinical course analyses. What we have shown here using an a priori hypothesis and testing in a prospective cohort may provide a method for explaining part of the "missing heritability" puzzle in MS. Nonetheless, it is essential

to replicate our results in other independent longitudinal cohorts to confirm the findings for future clinical use.

Therefore, in the post-GWAS world, the focus should shift from defining risk associations (using case-control study design in large GWAS) to defining genetic determinates of clinical course (using a priori hypotheses in longitudinal cohort study designs), either by assessing risk genes discovered by GWAS or by candidate gene studies utilizing GWAS genotype data. Chapters 4 and 5 are primarily based on the concepts that we propose here.

6.3 MBP and MS clinical course

Most MS patients present with a first clinical episode of demyelination (FDE), which has an uncertain prognosis and few generalizable markers of subsequent disease severity. Although these patients face an uncertain future, early intervention at this time may significantly alter future clinical outcomes. However, differentiating patients with good and poor prognoses is not yet possible. Identifying genetic and other risk factors that modulate early clinical course may assist the development of risk algorithms to better predict subsequent disease.

After an FDE, there are no clear genetic indicators of future conversion to MS, risk of relapse, or rate of disability accumulation. In Chapter 5, we developed an a priori hypothesis, considering the significant role of myelin basic protein (MBP) in CNS myelination, that genetic variants in the MBP gene may significantly affect MS clinical outcomes. Utilizing a longitudinal cohort study of persons who have had an FDE and who were followed up to a 5-year review, we found that the risk genotype of rs12959006 within *MBP* not only significantly predicted relapse (HR=1.74, 95%

CI=1.19-2.56, p=0.005) but also significantly predicted greater annualized ΔEDSS (β=0.18, 95% CI=0.06-0.30, p=0.004). Translating these findings into clinical outcomes for the participants followed for 5 years, those carrying the risk genotype of rs12959006 would have a 1.74-fold increased risk for relapse occurrence and an EDSS score 0.9 points greater compared with those carrying the non-risk genotype. We also found that the risk genotype significantly interacted with the known MS environmental risk factors to affect baseline anti-HHV6 IgG in predicting time to MS ($p_{interaction}$=0.05) and relapse ($p_{interaction}$=0.02), providing evidence that gene-environmental interaction can influence the clinical course of MS and thus explaining part of the missing MS heritability. The interactions provide evidence to support the molecular mimicry hypothesis, in which MBP-reactive T-cells may cross react with HHV6 epitopes and result in autoimmunity.

For the first time, we have demonstrated that genetic variants in biologically plausible MS candidate genes predicted worse MS clinical course using three key metrics. These results, if replicated, may aid in the development of prognostic algorithms in the early disease period of MS. Our results based on in silico analysis also provide direction for future studies to investigate the functional variations associated with the binding sites of miR-218 and miR-188-3p. Other research has already shown that in the white matter of MS patients, miR-218 expression was significantly down regulated compared with controls[18]. Therefore, studies focusing on the roles of these miRNAs in the post-transcriptional modification of the MBP gene may provide further insights into mechanisms affecting processes involved in myelination and demyelination.

6.4 miR-146a in predicting disease onset and relapse

miRNA-associated post-transcriptional gene silencing is an epigenetic control of gene expression that also may account for part of the missing heritability in MS[19]. Until now, extensive studies have only focused on the expression of miRNAs. No studies have investigated the function of variations within the miRNA coding genes and how such variations might predict disease onset and relapse.

In MS, the expression of miR-146a is known to be altered as a result of EBV infection[20] and to be upregulated in active MS brain lesions[21]. The only common variant within the miR-146a gene has been found to affect its expression. In Chapter 5, I assessed the effects of the only functional common variant in miR-146a using a longitudinal cohort study of persons who have had a FDE and who were followed up to the 5-year review.

In the univariate analysis, we found that the risk genotype (GC+CC) of rs2910164 significantly predicted relapse (HR=2.09 (95% CI: 1.42, 3.06), $p<0.001$). Upon further expansion to examine the gene-environmental interaction, we found a significant additive interaction between rs2910164 and EBV infection in predicting conversion to MS and relapse. The levels of anti-EBNA-1 and EBNA-2 IgG interacted with the risk genotype in predicting both MS and relapse, such that the combination of both the genetic and environmental risk factors predicted the outcome more strongly than the sum of each component alone. A similar additive interaction was observed for a history of infectious mononucleosis, but it only reached significance in predicting relapse. There was some evidence of a multiplicative interaction of similar directionality, particularly for relapse, but this interaction did

not reach statistical significance. While the test for multiplicative interaction is often applied for its ease, it has been argued that the assessment of interaction on the additive scale may be better correlated with biological interaction compared with a multiplicative interaction, thus making additive interaction more indicative of the underlying causal mechanism[22]. This demonstration of a significant interaction is germane to this study because it provides a supportive line of evidence for the direct association of the risk variant with clinical outcomes. In other words, by demonstrating that the environmental factors are more strongly associated with clinical outcomes among those carrying the risk variant in the manner expected by the known relationship between this miRNA and these environmental factors, the veracity of the main effects is substantiated.

Our findings support the hypothesis that we proposed in Chapter 1, that the complex interplay of environmental risk factors and epigenetic changes amplifies the risk of developing MS. This is one of the first studies to directly support this hypothesis and provide guidance for future studies aimed at identifying the role of epigenetics in the missing heritability of MS.

6.5 Final conclusion of PhD

This thesis covers a variety of topics using different study designs and methods to study the genetic variants and environmental factors influencing the onset and progression of MS, and it has revealed part of the missing heritability of MS. The key conclusions are as follows:

- There is shared polygenetic risk between EBNA-1 GWAS and MS GWAS; in addition to the main effects of the HLA region, the non-MHC region (1p22.1,

3p24.1, 3q13.33, and 10p15.1) also mediated the association of anti-EBNA-1 IgG titer with MS risk.

- Variations within *IL2RB* and in the 3' region of *TNFRSF1B* showed a significant interaction with TNF-α as a predictor of relapse, whereas variations in the 5' region of *IL3* and in the 3' region of *TNFRSF1B* showed a significant interaction between IFN-γ and relapse.
- People with an FDE who carry the risk SNP of rs12969006 in *MBP* have a worse outcome on all three measures of disease progression (conversion to MS, relapse rate, and EDSS progression), and the risk SNP of rs129569006 in *MBP* interacts with serological markers of prior HHV6 infection to predict clinical course post FDE.
- The only common variant within the miR-146a gene not only significantly predicts relapse in MS but also interacts with EBV infection in predicting MS and relapse.

6.6 Future directions

The QIMR twin families study and the AusLong study cohorts are a significant potential data source, with much more data that can be evaluated. Based on the results presented in this thesis, future work utilizing these datasets and/or collaborating with other research groups can be suggested as follows:

- The analysis of larger EBNA-1 GWAS datasets and/or the analysis of other markers of EBNA immune responses (e.g., EBNA-2 or EBNA-3 titers) should provide a novel opportunity to identify new and characterize existing genetic risk factors for MS.

- Functional studies focusing on miR-218 and miR-188-3p in the post-transcriptional modification of MBP and thus in the modification of myelination and demyelination could provide further mechanistic insights.
- A model that combines all the risk factors (genetic risk factors and environmental factors) should be developed to generate a score for each individual and determine how the score predicts clinical course. The ultimate aim of this model is for clinical prognosis, allowing doctors to determine the prognosis and treatment of MS based on this score.
- Lastly and importantly, the replication of our results from Chapter 3, 4 and 5 using other studies with larger sample sizes and longer follow-up is essential. However, due to the high cost and logistical limitations, there are presently no such studies available, thus supporting the importance of collaboration between multiple centers to allow these studies.

As a final conclusion, this thesis presents the condensation of over three years' work in which we sought to reveal the missing heritability puzzle of MS. Hopefully, as an ultimate goal of any research study, our work and the future work will be of benefit to MS patients.

6.7 References

1. van Dongen J, Slagboom PE, Draisma HH, Martin NG, Boomsma DI. The continuing value of twin studies in the omics era. *Nat Rev Genet.* 2012;13(9):640-653.

2. Lucas RM, Hughes AM, Lay ML, et al. Epstein-Barr virus and multiple sclerosis. *J. Neurol. Neurosurg. Psychiatry.* 2011;82(10):1142-1148.

3. Nyholt DR. SECA: SNP effect concordance analysis using genome-wide association summary results. *Bioinformatics.* 2014;30(14):2086-2088.

4. van Aalderen MC, Remmerswaal EB, Verstegen NJ, et al. Infection history determines the differentiation state of human CD8+ T cells. *J. Virol.* 2015;89(9):5110-5123.

5. Parnell GP, Gatt PN, Krupa M, et al. The autoimmune disease-associated transcription factors EOMES and TBX21 are dysregulated in multiple sclerosis and define a molecular subtype of disease. *Clin. Immunol.* 2014;151(1):16-24.

6. Burns A, Iliffe S. Alzheimer's disease. *BMJ.* 2009;338:b158.

7. Jankovic J. Parkinson's disease: clinical features and diagnosis. *J. Neurol. Neurosurg. Psychiatry.* 2008;79(4):368-376.

8. Xia Z, Chibnik LB, Glanz BI, et al. A putative Alzheimer's disease risk allele in PCK1 influences brain atrophy in multiple sclerosis. *PLoS One.* 2010;5(11):e14169.

9. Desikan RS, Schork AJ, Wang Y, et al. Genetic overlap between Alzheimer's disease and Parkinson's disease at the MAPT locus. *Mol. Psychiatry.* 2015;20(12):1588-1595.

10. Wang YQ, Tang BS, Yang Y, et al. Relationship between Alzheimer's disease GWAS-linked top hits and risk of Parkinson's disease with or without cognitive decline: a Chinese population-based study. *Neurobiol. Aging.* 2016;39:217 e219-211.

11. Zhu XC, Cao L, Tan MS, et al. Association of Parkinson's Disease GWAS-Linked Loci with Alzheimer's Disease in Han Chinese. *Mol. Neurobiol.* 2016.

12. International Multiple Sclerosis Genetics C, Bush WS, Sawcer SJ, et al. Evidence for polygenic susceptibility to multiple sclerosis--the shape of things to come. *Am. J. Hum. Genet.* 2010;86(4):621-625.

13. Navikas V, Link H. Review: cytokines and the pathogenesis of multiple sclerosis. *J. Neurosci. Res.* 1996;45(4):322-333.

14. Zhou Y, Taylor B, van der Mei I, et al. Genetic variation in PBMC-produced IFN-gamma and TNF-alpha associations with relapse in multiple sclerosis. *J. Neurol. Sci.* 2015;349(1-2):40-44.

15. Luo XJ, Li M, Huang L, et al. The interleukin 3 gene (IL3) contributes to human brain volume variation by regulating proliferation and survival of neural progenitors. *PLoS One.* 2012;7(11):e50375.

16. Cavanillas ML, Alcina A, Nunez C, et al. Polymorphisms in the IL2, IL2RA and IL2RB genes in multiple sclerosis risk. *Eur. J. Hum. Genet.* 2010;18(7):794-799.

17. Diogo D, Kurreeman F, Stahl EA, et al. Rare, low-frequency, and common variants in the protein-coding sequence of biological candidate genes from GWASs contribute to risk of rheumatoid arthritis. *Am. J. Hum. Genet.* 2013;92(1):15-27.

18. Noorbakhsh F, Ellestad KK, Maingat F, et al. Impaired neurosteroid synthesis in multiple sclerosis. *Brain.* 2011;134(Pt 9):2703-2721.

19. Zhou Y, Simpson S, Jr., Holloway AF, Charlesworth J, van der Mei I, Taylor BV. The potential role of epigenetic modifications in the heritability of multiple sclerosis. *Mult. Scler.* 2014;20(2):135-140.

20. Godshalk SE, Bhaduri-McIntosh S, Slack FJ. Epstein-Barr virus-mediated dysregulation of human microRNA expression. *Cell Cycle.* 2008;7(22):3595-3600.

21. Junker A, Krumbholz M, Eisele S, et al. MicroRNA profiling of multiple sclerosis lesions identifies modulators of the regulatory protein CD47. *Brain.* 2009;132(Pt 12):3342-3352.

22. Rothman KJ. Synergy and antagonism in cause-effect relationships. *Am. J. Epidemiol.* 1974;99(6):385-388.

Appendix: Other Publications during PhD (My contributions include: data analysis and paper revising).

Zochling J, Newell F, Charlesworth JC, Leo P, Stankovich J, Cortes A, Zhou Y, et al. (2014) An Immunochip-based interrogation of scleroderma susceptibility variants identifies a novel association at DNASE1L3. Arthritis Res Ther 16 (5):438. doi:10.1186/s13075-014-0438-8.

Zochling et al. Arthritis Research & Therapy 2014, 16:438
http://arthritis-research.com/content/16/5/438

arthritis research&therapy

RESEARCH ARTICLE Open Access

An Immunochip-based interrogation of scleroderma susceptibility variants identifies a novel association at *DNASE1L3*

Jane Zochling[1], Felicity Newell[2], Jac C Charlesworth[1], Paul Leo[2], Jim Stankovich[1], Adrian Cortes[2], Yuan Zhou[1], Wendy Stevens[3], Joanne Sahhar[4], Janet Roddy[5], Peter Nash[6], Kathleen Tymms[7], Maureen Rischmueller[8], Sue Lester[8], Susanna Proudman[9] and Matthew A Brown[2*]

Abstract

Introduction: The aim of the study was to interrogate the genetic architecture and autoimmune pleiotropy of scleroderma susceptibility in the Australian population.

Methods: We genotyped individuals from a well-characterized cohort of Australian scleroderma patients with the Immunochip, a custom array enriched for single nucleotide polymorphisms (SNPs) at immune loci. Controls were taken from the 1958 British Birth Cohort. After data cleaning and adjusting for population stratification the final dataset consisted of 486 cases, 4,458 controls and 146,525 SNPs. Association analyses were conducted using logistic regression in PLINK. A replication study was performed using 833 cases and 1,938 controls.

Results: A total of eight loci with suggestive association ($P < 10^{-4.5}$) were identified, of which five showed significant association in the replication cohort (*HLA-DRB1, DNASE1L3, STAT4, TNP03-IRF5* and *VCAM1*). The most notable findings were at the *DNASE1L3* locus, previously associated with systemic lupus erythematosus, and *VCAM1*, a locus not previously associated with human disease. This study identified a likely functional variant influencing scleroderma susceptibility at the *DNASE1L3* locus; a missense polymorphism rs35677470 in *DNASE1L3*, with an odds ratio of 2.35 ($P = 2.3 \times 10^{-10}$) in anti-centromere antibody (ACA) positive cases.

Conclusions: This pilot study has confirmed previously reported scleroderma associations, revealed further genetic overlap between scleroderma and systemic lupus erythematosus, and identified a putative novel scleroderma susceptibility locus.

Introduction

Systemic sclerosis (SSc) or scleroderma is a multisystem, autoimmune disorder characterised by progressive vascular, inflammatory and fibrotic dysfunction. Skin and visceral complications of cardiac, pulmonary, gastrointestinal, muscle and renal disease can have devastating effects on quality of life and life expectancy [1].

Scleroderma has a well-established genetic component [2-4]. Most of the identified SSc susceptibility loci overlap with those of other autoimmune diseases, in particular the rheumatic disorders such as rheumatoid arthritis and systemic lupus erythematosus (SLE) [5]. For example, Carmona *et al.* recently confirmed an association between SSc and the SLE risk haplotype at the *IRF5* locus [6] and Martin *et al.* recently performed a pan-meta-analysis of SSc and SLE to look at the susceptibility overlap between the two diseases [7]. Using 6,835 cases and 14,274 controls they identified a novel pleiotropic locus at *KIAA0319L* on chromosome 1 and identified two SLE loci (near *PXK* and *JAZF1*) that also contribute to SSc [7].

To identify further susceptibility loci and to explore the genetic overlap with other antibody-mediated immune diseases, we undertook an SSc association study using the Immunochip, a custom array including SNPs of interest in a wide variety of autoimmune disorders [8]. A replication study was then performed using previously published case

* Correspondence: Matt.Brown@uq.edu.au
[2]University of Queensland Diamantina Institute, Translational Research Institute, Princess Alexandra Hospital, 37 Kent Street, Woolloongabba, Brisbane, QLD 4102, Australia
Full list of author information is available at the end of the article

data from a genome-wide association study (GWAS) of scleroderma [9], and control data from a GWAS of bone density variation [10].

Methods

Samples

We selected 532 cases for genotyping from the Australian Scleroderma Cohort Study (ASCS) [11]; a prospective study of risk factors for clinically important outcomes in SSc. They fulfilled either the American College of Rheumatology (ACR) criteria for classification of SSc [12] or the Medsger criteria for limited SSc (to enable broad representation of the disease spectrum) [13]. The study was approved by the Human Research Ethics Committee (Tasmanian) Network and human research ethics committees of St. Vincent's Hospital and Monash Medical Centre, Melbourne, VIC; Sunshine Coast Rheumatology, Maroochydore, QLD; Royal Adelaide Hospital, Adelaide, SA; St George Hospital, Sydney, NSW; Royal Perth Hospital, Perth, WA; and Prince Charles Hospital, Brisbane, QLD. All patients gave written, informed consent.

Genotyping

Cases were genotyped with the Immunochip, an Illumina Infinium SNP microarray (Illumina Inc., San Diego, CA, USA) [8], at the University of Queensland Diamantina Institute, Brisbane, QLD, Australia. The Immunochip contains 195,806 common and rare SNPs of interest in a wide variety of autoimmune disorders. Control genotypes were obtained from 4,537 samples from the 1958 British Birth Cohort [14]. Genotypes were called using the Illumina GenTrain clustering algorithm. Cases and controls were clustered separately.

For replication purposes, genotypes from 833 SSc cases were obtained from dbGAP (dbGaP Study Accession: phs000357.v1.p1) [9,15]. These samples were genotyped with the Illumina Human610-Quad v1.0 BeadChip. Control genotypes were obtained from 1,938 subjects of white British ancestry genotyped as part of the Anglo-Australasian Osteoporosis Genetics Consortium program using either Illumina Infinium II HumHap300 or 370CNV chips [10].

Statistical analyses

Genotype data were analysed with PLINK [16] and R [17]. There were no duplicate or closely related cases. Case (n = 2) and control (n = 3) samples with call rates less than 90% were excluded. SNPs were excluded based on Hardy-Weinberg disequilibrium ($P < 10^{-6}$), call rates less than 90%, fewer than two occurrences of the minor allele, and significantly different rates of missingness ($P < 10^{-4}$) between cases and controls. Eigenstrat [18] was run on a pruned SNP set with default settings to exclude population ancestry outliers and ensure cases and controls were ethnically matched. Subjects lying more than six standard deviations from the mean of any principal component were excluded (Immunochip set 44 cases and 76 controls excluded; replication set 129 cases and 31 controls excluded).

Four-digit classical MHC allele dosages at *HLA-A, HLA-B, HLA-C, HLA-DRB1, HLA-DQA1* and *HLA-DQB1* were imputed using HLA*IMP [19,20] in the Immunochip dataset, and tested for association.

Logistic regression was used for all association analyses using principal components derived from the Eigenstrat analysis as covariates to control for population stratification. A single principal component was used for the Immunochip analysis and two principal components for the replication dataset; including further principal components for either set did not reduce the genomic inflation factor further. A negative control set of 2,805 SNPs outside the major histocompatibility complex (MHC), associated with reading and learning, schizophrenia and psychosis [19], was used to estimate the genomic inflation factor and calculate adjusted *P* values. Genotype intensity cluster plots were manually examined for all suggestive associations with unadjusted *P* values less than $10^{-4.5}$, a threshold used in previous GWAS [21]. To test for secondary association signals at each locus, genotypes at the most significant variant were added to the logistic regression model as a covariate, and all other variants at the locus were tested. To correct secondary analyses for multiple testing, we estimated the effective number of independent tests at loci using the eigenvalues of the matrix of correlations between SNPs [22], as implemented in SNPSpD [23].

For variants associated with SSc at the nominally suggestive threshold ($P < 10^{-4.5}$), we also tested for differences in allele frequencies between patients with different disease subtypes and MHC genotypes.

For replication, cases and controls genotypes were imputed as implemented in IMPUTE2 [24] with the use of the merged 1000 Genomes and UK10K reference dataset. All SNPs we were attempting to replicate had an info score of >0.7. Meta-analysis was performed using METAL weighted by inverse variance [25]. Power was calculated using the Genetic Power Calculator [26].

Results

After sample and SNP exclusions, in the Immunochip dataset genotypes were analysed at 145,921 autosomal SNPs in 486 cases and 4,458 controls, with a genomic inflation factor of 1.02 (Q-Q plot, Figure S1 in Additional file 1). Eighty-six percent of the cases were female, mean age was 60 years and mean disease duration 14 years. Twenty-five percent of cases had diffuse disease, 72% limited pattern, and 3% were intermediate; 43% were anti-centromere antibody (ACA) positive and 15% anti-Scl-70 antibody positive. Considering the replication set, 700 scleroderma cases and 1,899 controls remained after

quality control. Of the replication cases, 36% had diffuse disease, 64% limited pattern; 32% were ACA positive. The genomic inflation factor for the replication set was 1.045. The Immunochip study has 80% power to detect associations at $P < 10^{-4.5}$ for a variant with minor allele frequency (MAF) of 0.3 with D' = 0.8 with a SNP with heterozygote odds ratio (OR) of 1.65.

We detected suggestive associations in the Immunochip data (uncorrected $P < 10^{-4.5}$) at eight loci (Table 1), five of which showed association in the replication cohort (*HLA-DRB1*, *DNASE1L3*, *STAT4*, *TNPO3-IRF5*, and *VCAM1*) (Table 1, Manhattan plot Figure 1). There was also evidence of association at the previously reported genome-wide significant *CD247* SNP [9] (rs2056626, OR = 0.76, $P = 1.1 \times 10^{-4}$), but no evidence of association at the genome-wide significant *TNIP1* locus [27] (rs2233287, $P = 0.94$). The overlap between previously reported genome-wide significant SSc loci (outside the MHC) and our data, at an unadjusted $P < 0.01$, is shown in Table 2. Not all previously associated SSc loci could be investigated owing to the limitations of markers available on the array.

The strongest SNP association (Table 1) was with rs2857130 in the MHC. Testing imputed MHC alleles for association, the strongest signal was for *DRB1*11:04* (Table 3). There was also a protective association with *DRB1*07:01*, but no other alleles or SNPs showed evidence of association after conditioning on these two. In particular, there was little evidence of association with the rare *DRB1*11:03* allele, which only differs from *DRB1*11:04* at one site encoding part of a hypervariable, peptide-binding region (R71E; OR 1.46, $P = 0.35$), and a model with *DRB1*11:03* and *DRB1*11:04* dosage combined does not fit as well as a model with *DRB1*11:04* dosage alone.

Consistent with previous findings, *DRB1*11:04* is a particularly strong risk factor for Scl70-positive SSc [28] (Table 3). However, compared to controls the frequency of this allele is also elevated in ACA-positive cases and in cases negative for both antibodies. The protective effect of *DRB1*07:01* is strongest against ACA-positive disease, and there was no evidence that this allele protects against anti-Scl70 antibody-positive disease.

The most significant *STAT4* locus association was at the SNP rs13426947 (Table 1). This was one of 15 associated intronic SNPs in *STAT4* ($P < 3 \times 10^{-4}$), all correlated with rs13426947 ($r^2 > 0.55$, MAF 0.19 to 0.26). These included the top-ranked SNP at this locus from a previous SSc GWAS (rs3821236, $P = 2.6 \times 10^{-5}$, $r^2 = 0.97$ with rs13426947) [9]. No SNPs in this region were significantly associated with SSc ($P < 0.01$) after conditioning on rs13426947. rs13426947 also showed association in the replication set ($P = 5.4 \times 10^{-3}$), and in the combined analysis ($P = 6.1 \times 10^{-7}$).

At the *IRF5/TNPO3* locus, there was a group of 23 associated SNPs ($P < 5 \times 10^{-5}$) highly correlated with the top-ranked SNP rs34381587 ($r^2 > 0.90$, MAF 0.11 to 0.13), including the top-ranked SNP reported previously (rs10488631, $P = 2.4 \times 10^{-5}$, $r^2 = 0.998$ with rs34381587) [9]. There was some evidence of association with a rare SNP between the two genes (128,379,270 base pairs,

Table 1 SNPs associated with SSc ($P < 10^{-4.5}$) in an analysis of Immunochip genotypes for 486 cases and 4,458 controls, replication cohort of 700 SSc cases (220 ACA positive) and 1,889 controls, and combined

Chr	Location range (base pairs, hg18)	Best SNP	MAF cases	MAF controls	Immunochip OR	Immunochip *P* value	Replication OR	Replication *P* value	Combined analysis	Gene annotation[#]
6	32,039,116-32,888,448	rs2857130	0.293	0.389	0.68	2.8×10^{-7}	0.81	0.003	1.3×10^{-8}	Intergenic (HLA region)
1	101,009,225	Novel SNP	0.024	0.008	3.31	1.8×10^{-6}	2.49	0.031	1.9×10^{-7}	Intergenic (*VCAM1*)
3	58,158,676-58,289,303	rs35677470	0.122	0.083	1.63	3.4×10^{-6}	1.27	0.027	1.2×10^{-6}	*DNASE1L3* (coding)
3	58,158,676-58,289,303	rs35677470	0.174	0.083	2.36	2.3×10^{-10}	1.74	3.0×10^{-4}	8.71×10^{-13}	*DNASE1L3* (coding)[^]
7	128,372,852-128,499,110	rs34381587	0.158	0.113	1.53	1.2×10^{-5}	1.38	8.7×10^{-4}	5.2×10^{-8}	*IRF5/TNPO3* (intron)
16	73,863,956-74,046,823	rs11149824	0.469	0.391	1.35	1.4×10^{-5}	1.054	0.442	3.2×10^{-4}	*CFDP1* (intron)
2	43,775,459-43,784,213	rs13403030	0.384	0.318	1.36	1.5×10^{-5}	0.92	0.242	0.027	*PLEKHH2* (intron)
2	191,608,694-191,641,499	rs13426947	0.253	0.191	1.42	1.8×10^{-5}	1.25	5.4×10^{-3}	6.1×10^{-7}	*STAT4* (intron)
1	61,883,642	rs2886326	0.249	0.195	1.41	2.0×10^{-5}	0.96	0.606	7.4×10^{-3}	Intergenic (*TM2D1*)

[#]Gene annotation is based on the location of the most significant SNP; [^]ACA-positive cases only. SNP, single nucleotide polymorphism; SSc, systemic sclerosis; ACA, anti-centromere antibody; Chr, chromosome; MAF, minor allele frequency; OR, odds ratio.

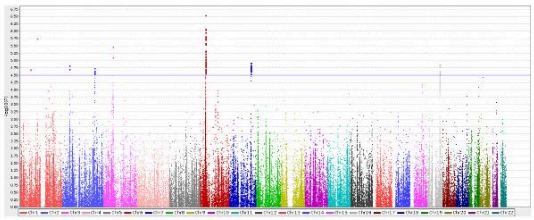


Figure 1 Manhattan Plot for association study finding from Immunochip genotyped cases and controls.

MAF 0.028, OR 1.83, $P = 6.0 \times 10^{-4}$; $P = 0.05$ after correction for the 83 independent tests). rs34381587 also showed association in the replication set ($P = 8.7 \times 10^{-4}$), and near genome-wide significance in the combined analysis ($P = 5.2 \times 10^{-8}$).

Association was observed and replicated at chromosome 3p14, spanning *DNASE1L3* to *AXOX2* (including *PXK*; Figure 2), a region that has previously been associated with both SLE [29] and SSc [7]. While the peak association with SLE was originally identified at an intronic SNP rs6445975 in *PXK* (a PX domain containing serine/threonine kinase), the strongest association with SSc on the Immunochip was at a missense SNP rs35677470 (R206C) in *DNASE1L3* (deoxyribonuclease I-like 3); 187 kb distal of rs6445975. Linkage disequilibrium between the two SNPs is modest ($r^2 = 0.13$), and there is weak evidence of a secondary association with rs6445975 in our data after conditioning on rs35677470 (OR = 1.19, $P = 0.03$). No other associations at this locus are significant after correction for multiple testing. The association with rs35677470 is confined to ACA-positive cases (estimated OR 2.36, $P = 2.3 \times 10^{-10}$), with no association in ACA-negative cases ($P = 0.76$). rs35677470 also showed association in the replication set ($P = 0.027$), and in the combined analysis ($P = 1.2 \times 10^{-6}$). As in the Immunochip analysis, in the replication set association was much stronger in the ACA-positive group ($P = 3.0 \times 10^{-4}$), and overall ($P = 8.71 \times 10^{-13}$). Association was particularly significant in limited scleroderma ($P = 3.36 \times 10^{-9}$ in overall dataset) compared with diffuse scleroderma ($P = 0.57$), consistent with the association of ACA antibody status with limited disease. The non-synonymous *DNASE1L3* variant is predicted to be deleterious to the protein product using *in silico* functional prediction tools including both SIFT [30] and PolyPhen [31]. Apart from the *DNASE1L3* locus, no other associations showed evidence of heterogeneity by antibody status, or between limited and diffuse SSc.

A novel intergenic SNP in *VCAM1* achieved suggestive association with SSc overall in Immunochip cases (OR = 3.31, $P = 1.8 \times 10^{-6}$). Association was seen with both limited ($P = 1.8 \times 10^{-4}$), and diffuse disease (8.9×10^{-5}). The findings for overall and limited disease were supported in the replication dataset ($P = 0.031$ and 3.1×10^{-3} respectively) but not with diffuse disease ($P = 0.7$). A combined meta-analysis gave P values of 1.9×10^{-7} for all cases, 1.9×10^{-6} for limited and 4.3×10^{-4} for diffuse disease.

Discussion

The major novel finding of this study is the significant association of a functional SNP (rs35677470) in *DNASE1L3*

Table 2 Association finding for previously reported loci achieving *P* <0.01 in Immunochip, with findings in replication and overall datasets

SNP	Chr	Location (hg18; bp)	Locus	MAF cases	MAF controls	OR	Unadjusted *P* value	Replication OR	Replication *P* value	Combined analysis	References
rs2056626	1	165,687,049	*CD247*	0.359	0.428	0.76	1.1×10^{-4}	0.84	0.007	3.9×10^{-6}	[6,7,11]
rs3821236	2	191,611,003	*STAT4*	0.255	0.194	1.41	2.6×10^{-5}	1.24	0.008	1.2×10^{-6}	[6,7]
rs10168266	2	191,644,049	*STAT4*	0.249	0.190	1.40	4.1×10^{-5}	1.30	0.001	1.9×10^{-7}	[11]
rs7574865	2	191,672,878	*STAT4*	0.286	0.225	1.35	1.2×10^{-4}	1.15	0.066	5.7×10^{-5}	[7]
rs2176082	3	58,306,226	*PXK*	0.343	0.296	1.31	2.6×10^{-4}	1.15	0.046	7.2×10^{-5}	[16]
rs4728142	7	128,361,203	*TNP03-IRF5*	0.497	0.449	1.23	3.2×10^{-3}	1.23	0.002	2.1×10^{-5}	[6,11]
rs10488631	7	128,381,419	*TNP03-IRF5*	0.156	0.114	1.50	2.4×10^{-5}	1.36	0.001	1.6×10^{-7}	[6,7,11]
rs12531711	7	128,404,702	*TNP03-IRF5*	0.156	0.113	1.51	2.1×10^{-5}	1.38	0.001	8.6×10^{-8}	[11]
rs1378942	15	72,864,420	*CSK*	0.379	0.319	1.27	8.7×10^{-4}	1.17	0.023	7.8×10^{-5}	[11]

Chr chromosome; bp, base pairs; MAF, minor allele frequency; OR, odds ratio.

Table 3 Association testing results for MHC alleles *DRB1*11:04* and *DRB1*07:01*

	Results for *DRB1*11:04*			Results for *DRB1*07:01*		
Group	**Mean allele dosage**	**OR (95% CI)**	***P* value**	**Mean allele dosage**	**OR (95% CI)**	***P* value**
Controls (n =4458)	0.037	1 (ref)		0.287	1 (ref)	
Cases (n =486)	0.105	3.07 (2.00 – 4.71)	2.8×10^{-7}	0.167	0.58 (0.46 – 0.74)	1.4×10^{-5}
ACA-positive (44% of cases)	0.080	2.46 (1.28 – 4.73)		0.105	0.34 (0.22 – 0.53)	
ACA-negative (56% of cases)	0.118	3.62 (2.13 – 6.16)	0.19*	0.213	0.79 (0.59 – 1.07)	0.002*
Scl70-positive (16% of cases)	0.174	8.22 (3.74 – 18.1)		0.294	1.28 (0.79 – 2.06)	
Scl70-negative (84% of cases)	0.088	2.47 (1.52 – 4.02)	0.021*	0.145	0.49 (0.37 – 0.65)	0.0012*
ACA & Scl70-negative (40% of cases)	0.095	2.61 (1.36 – 5.00)		0.144	0.66 (0.46 – 0.96)	

*These *P* values are tests of heterogeneity, comparing allele dosages in cases positive and negative for the two antibodies. Mean dosage = mean expected number of copies of alleles carried by individuals in group. OR (odds ratio) = increase in odds of being a case for each 1-unit increase in allele dosage. MHC, major histocompatibility complex; CI, confidence interval; ACA, anti-centromere antibody.

with ACA-positive SSc ($P = 2.3 \times 10^{-10}$). This locus has been previously reported in a SLE GWAS [29] and meta-analysis [32], however the peak associations were reported for SNPs in the nearby gene *PXK*. While the *PXK* and *DNASE1L3* associations may be independent, no SNPs on the Illumina HumanHap300 chip used in the SLE GWAS are good tags for the missense SNP in *DNASE1L3* (maximum r^2 of 0.21). More recently, a functional variant in *DNASE1L3* was implicated in a familial form of SLE [33], and a pan-meta-analysis of SLE and SSc also confirmed the locus near *PXK*, but in particular for ACA-positive SSc (rs2176082 [7]; $P = 1.4 \times 10^{-4}$ in our data) strengthening the evidence that this locus plays a role in both diseases. These findings were independently identified in a study published whilst the current manuscript was in review [34].

The associated missense *DNASE1L3* variant in our data, (rs35677470 encoding R206C) affects a highly conserved residue and there is very strong evidence that this results in loss of function of the protein [35]. *DNASE1L3* encodes a member of the DNase family and functions as an endonuclease capable of cleaving DNA, mediating the breakdown of DNA during apoptosis. Al-Mayouf *et al.* [33] hypothesised that, in the context of SLE, dysfunction of this gene may lead to impaired DNA breakdown and clearance from apoptotic cells, resulting in the formation of self-directed DNA-specific antibodies and immune complexes. Since the same kinds of DNA-driven immune complexes (such as anti-nuclear and ACA antibodies) are also characteristic of SSc, this hypothesis is also applicable.

Suggestive association was also observed between a novel SNP in *VCAM1* and overall scleroderma and in limited disease cases, both of which were also associated in the replication dataset. *VCAM1* has not previously been reported to be associated with scleroderma or SLE. VCAM-1 is a member of the Ig superfamily and encodes a cell surface sialoglycoprotein expressed by cytokine-activated endothelium. It mediates leukocyte-endothelial cell adhesion and

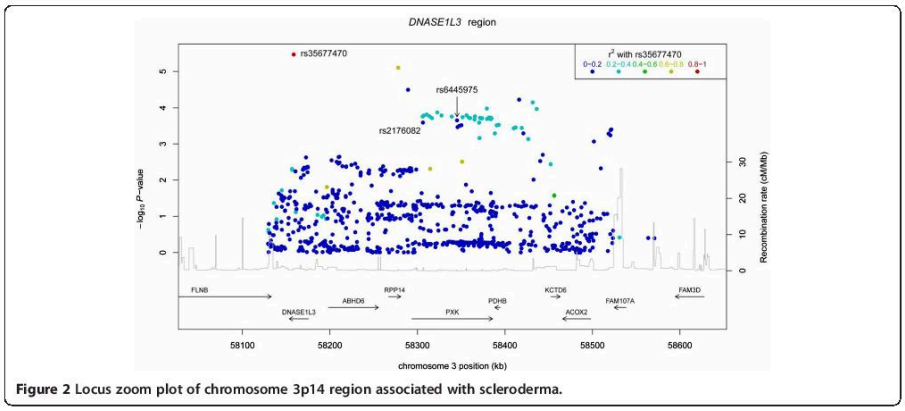


Figure 2 Locus zoom plot of chromosome 3p14 region associated with scleroderma.

signal transduction. VCAM-1 levels have previously been shown to be elevated in early, inflammatory-phase scleroderma [36] and in limited scleroderma [37].

There were no stand-out functional variants at the *STAT4* and *IRF5/TNPO3* loci. These associations included many highly correlated SNPs, indicating that larger sample sizes and/or functional studies will be needed to understand and dissect these associations. It would, however, be prudent to investigate any functional variation at these loci identified in related autoimmune diseases, particularly SLE. The other novel associations are merely suggestive and require confirmation in additional datasets.

Conclusions

There is a significant association of a functional SNP in *DNASE1L3* with anti-centromere antibody-positive SSc, previously reported in SLE. There is strong evidence for a loss of function of the protein. A novel association was also observed and replicated with an intergenic SNP in *VCAM1*.

This study serves to highlight that, even with a small but well-characterised disease cohort, significant associations can be obtained by tools such as the Immunochip, which are targeted towards analysis of disease-relevant and occasionally functional variation.

Additional file

Additional file 1: Figure S1. Q-Q plot considering reading and learning disability, psychosis and schizophrenia SNPs, excluding the MHC.

Abbreviations

ACA: anti-centromere antibody; bp: base pair; Chr: chromosome; CI: confidence interval; GWAS: genome-wide association study; MAF: minor allele frequency; MHC: major histocompatibility complex; OR: odds ratio; SLE: systemic lupus erythematosus; SNP: single nucleotide polymorphism; SSc: systemic sclerosis.

Competing interests

The authors declare that they have no competing interests. There was no financial support or other benefits from commercial sources for the work reported on in the manuscript that could create a potential conflict of interest or the appearance of a conflict of interest with regard to the work.

Authors' contributions

JZ participated in the design of the study, provided patient data for the analysis and helped to draft the manuscript. FN, JC and PL performed the statistical analysis and helped to draft the manuscript. JSt, AC and YZ assisted in performing the statistical analysis. WS, JSa, JR, PN, KT and MR participated in the design of the study and provided patient data for the analysis. SL assisted in study design and performed the DNA extraction. SP participated in the design of the study and provided patient data for the analysis. MB participated in the design and coordination of the study, assisted in the genotyping, provided expert advice regarding the statistical analysis, and drafted the manuscript. All authors assisted in data interpretation, and read and approved the final manuscript.

Acknowledgments

The Wellcome Trust Case Control Consortium, Australian Scleroderma Interest Group, Actelion Australia, Bayer, CSL Biotherapies, GlaxoSmithKline Australia, Pfizer, NHMRC Program Grant 569938 and the Australian Cancer Research Foundation - Tasmanian Inherited Cancer Centre Computing Facility. MB was supported by an NHMRC Senior Principal Research Fellowship APP1024879.
The data used for the replication analyses described in this paper were obtained from the database of Genotypes and Phenotypes (dbGaP), at http://www.ncbi.nlm.nih.gov/gap. Genotype and phenotype data for the genome-wide association study in systemic sclerosis study (dbGaP accession number phs0 00 357 .v1.p1) were provided by Maureen D. Mayes, University of Texas genome-wide association study in systemic sclerosis Page 5 of 7 Health Science Center, Houston, Texas. Funding support for the original study was provided by the National Institutes of Health, and by other sources detailed in *Nat Genet* 2010 May; 42: 426–9. Genome-wide association study of systemic sclerosis identifies CD247 as a new susceptibility locus. Radstake TR, Gorlova O, Rueda B, Martin JE, Alizadeh BZ, Palomino-Morales R, Coenen MJ, Vonk MC, Voskuyl AE, Schuerwegh AJ, Broen JC, van Riel PL, van 't Slot R, Italiaander A, Ophoff RA, Riemekasten G, Hunzelmann N, Simeon CP, Ortego-Centeno N, González-Gay MA, González-Escribano MF; Spanish Scleroderma Group, Airo P, van Laar J, Herrick A, Worthington J, Hesselstrand R, Smith V, de Keyser F, Houssiau F, Chee MM, Madhok R, Shiels P, Westhovens R, Kreuter A, Kiener H, de Baere E, Witte T, Padykov L, Klareskog L, Beretta L, Scorza R, Lie BA, Hoffmann-Vold AM, Carreira P, Varga J, Hinchcliff M, Gregersen PK, Lee AT, Ying J, Han Y, Weng SF, Amos CI, Wigley FM, Hummers L, Nelson JL, Agarwal SK, Assassi S, Gourh P, Tan FK, Koeleman BP, Arnett FC, Martin J, Mayes MD.

Author details

[1]Menzies Research Institute Tasmania, University of Tasmania, 17 Liverpool Street, Hobart, TAS 7000, Australia. [2]University of Queensland Diamantina Institute, Translational Research Institute, Princess Alexandra Hospital, 37 Kent Street, Woolloongabba, Brisbane, QLD 4102, Australia. [3]Rheumatology Department, St Vincent's Hospital, 41 Victoria Parade, Fitzroy, Melbourne, VIC 3065, Australia. [4]Department of Rheumatology, Monash Medical Centre, 246 Clayton Road, Clayton, Melbourne, VIC 3168, Australia. [5]Department of Rheumatology, Royal Perth Hospital, 197 Wellington Street, Perth, WA 6000, Australia. [6]Research Unit, Sunshine Coast Rheumatology, Denna Street, Maroochydore, QLD 4558, Australia. [7]Canberra Rheumatology, 40 Marcus Clarke Street, Canberra, ACT 2601, Australia. [8]Department of Rheumatology, The Queen Elizabeth Hospital, 28 Woodville Road, Woodville, SA 5011, Australia. [9]Rheumatology Unit, Royal Adelaide Hospital, North Terrace, Adelaide, SA 5000, Australia.

Received: 29 October 2013 Accepted: 26 August 2014
Published online: 21 October 2014

References

1. Steen VD, Medsger TA Jr: **Severe organ involvement in systemic sclerosis with diffuse scleroderma.** *Arthritis Rheum* 2000, **43:**2437–2444.
2. Romano E, Manetti M, Guiducci S, Ceccarelli C, Allanore Y, Matucci-Cerinic M: **The genetics of systemic sclerosis: an update.** *Clin Exp Rheumatol* 2011, **29:**S75–S86.
3. Englert H, Small-McMahon J, Chambers P, O'Connor H, Davis K, Manolios N, White R, Dracos G, Brooks P: **Familial risk estimation in systemic sclerosis.** *Aust N Z J Med* 1999, **29:**36–41.
4. Arnett FC, Cho M, Chatterjee S, Aguilar MB, Reveille JD, Mayes MD: **Familial occurrence frequencies and relative risks for systemic sclerosis (scleroderma) in three United States cohorts.** *Arthritis Rheum* 2001, **44:**1359–1362.
5. Martin JE, Bossini-Castillo L, Martin J: **Unraveling the genetic component of systemic sclerosis.** *Hum Genet* 2012, **131:**1023–1037.
6. Carmona FD, Martin JE, Beretta L, Simeon CP, Carreira PE, Callejas JL, Fernandez-Castro M, Saez-Comet L, Beltran E, Camps MT, Egurbide MV, Airo P, Scorza R, Lunardi C, Hunzelmann N, Riemekasten G, Witte T, Kreuter A, Distler JH, Madhok R, Shiels P, van Laar JM, Fonseca C, Denton C, Herrick A, Worthington J, Schuerwegh AJ, Vonk MC, Voskuyl AE, Radstake TR, *et al*: **The systemic lupus erythematosus IRF5 risk haplotype is associated with systemic sclerosis.** *PLoS One* 2013, **8:**e54419.
7. Martin JE, Assassi S, Diaz-Gallo LM, Broen JC, Simeon CP, Castellvi I, Vicente-Rabaneda E, Fonollosa V, Ortego-Centeno N, Gonzalez-Gay MA, Espinosa G, Carreira P, Spanish Scleroderma Group; SLEGEN consortium; U.S. Scleroderma GWAS group; BIOLUPUS, Camps M, Sabio JM, D'alfonso S, Vonk MC, Voskuyl AE, Schuerwegh AJ, Kreuter A, Witte T, Riemekasten G, Hunzelmann N, Airo P,

Beretta L, Scorza R, Lunardi C, Van Laar J, Chee MM, Worthington J, *et al*: **A systemic sclerosis and systemic lupus erythematosus pan-meta-GWAS reveals new shared susceptibility loci.** *Hum Mol Genet* 2013, **22:**4021–4029.
8. Cortes A, Brown MA: **Promise and pitfalls of the Immunochip.** *Arthritis Res Ther* 2011, **13:**101.
9. Radstake TR, Gorlova O, Rueda B, Martin JE, Alizadeh BZ, Palomino-Morales R, Coenen MJ, Vonk MC, Voskuyl AE, Schuerwegh AJ, Broen JC, van Riel PL, van't Slot R, Italiaander A, Ophoff RA, Riemekasten G, Hunzelmann N, Simeon CP, Ortego-Centeno N, Gonzalez-Gay MA, Gonzalez-Escribano MF, Airo P, van Laar J, Herrick A, Worthington J, Hesselstrand R, Smith V, DeKeyser F, Houssiau F, Chee MM, *et al*: **Genome-wide association study of systemic sclerosis identifies CD247 as a new susceptibility locus.** *Nat Genet* 2010, **42:**426–429.
10. Duncan EL, Danoy P, Kemp JP, Leo PJ, McCloskey E, Nicholson GC, Eastell R, Prince RL, Eisman JA, Jones G, Sambrook PN, Reid IR, Dennison EM, Wark J, Richards JB, Uitterlinden AG, Spector TD, Esapa C, Cox RD, Brown SD, Thakker RV, Addison KA, Bradbury LA, Center JR, Cooper C, Cremin C, Estrada K, Felsenberg D, Gluer CC, Hadler J, *et al*: **Genome-wide association study using extreme truncate selection identifies novel genes affecting bone mineral density and fracture risk.** *PLoS Genet* 2011, **7:**e1001372.
11. Nikpour M, Hissaria P, Byron J, Sahhar J, Micallef M, Paspaliaris W, Roddy J, Nash P, Sturgess A, Proudman S, Stevens W: **Prevalence, correlates and clinical usefulness of antibodies to RNA polymerase III in systemic sclerosis: a cross-sectional analysis of data from an Australian cohort.** *Arthritis Res Ther* 2011, **13:**R211.
12. Masi AT, Subcommittee for Scleroderma Criteria of the American Rheumatism Association Diagnostic and Therapeutic Criteria Committee: **Preliminary criteria for the classification of systemic sclerosis (scleroderma).** *Arthritis Rheum* 1980, **23:**581–590.
13. LeRoy EC, Medsger TA Jr: **Criteria for the classification of early systemic sclerosis.** *J Rheumatol* 2001, **28:**1573–1576.
14. Evans DM, Spencer CC, Pointon JJ, Su Z, Harvey D, Kochan G, Oppermann U, Dilthey A, Pirinen M, Stone MA, Appleton L, Moutsianas L, Leslie S, Wordsworth T, Kenna TJ, Karaderi T, Thomas GP, Ward MM, Weisman MH, Farrar C, Bradbury LA, Danoy P, Inman RD, Maksymowych W, Gladman D, Rahman P, Morgan A, Marzo-Ortega H, Bowness P, Gaffney K, *et al*: **Interaction between ERAP1 and HLA-B27 in ankylosing spondylitis implicates peptide handling in the mechanism for HLA-B27 in disease susceptibility.** *Nat Genet* 2011, **43:**761–767.
15. **dbGAP** [http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000357.v1.p1]
16. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC: **PLINK: a tool set for whole-genome association and population-based linkage analyses.** *Am J Hum Genet* 2007, **81:**559–575.
17. **Bioconductor Project** [www.R-project.org]
18. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D: **Principal components analysis corrects for stratification in genome-wide association studies.** *Nat Genet* 2006, **38:**904–909.
19. Dilthey AT, Moutsianas L, Leslie S, McVean G: **HLA*IMP–an integrated framework for imputing classical HLA alleles from SNP genotypes.** *Bioinformatics* 2011, **27:**968–972.
20. Leslie S, Donnelly P, McVean G: **A statistical method for predicting classical HLA alleles from SNP data.** *Am J Hum Genet* 2008, **82:**48–56.
21. Sawcer S, Hellenthal G, Pirinen M, Spencer CC, Patsopoulos NA, Moutsianas L, Dilthey A, Su Z, Freeman C, Hunt SE, Edkins S, Gray E, Booth DR, Potter SC, Goris A, Band G, Oturai AB, Strange A, Saarela J, Bellenguez C, Fontaine B, Gillman M, Hemmer B, Gwilliam R, Zipp F, Jayakumar A, Martin R, Leslie S, Hawkins S, Giannoulatou E, *et al*: **Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis.** *Nature* 2011, **476:**214–219.
22. Li J, Ji L: **Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix.** *Heredity* 2005, **95:**221–227.
23. Nyholt DR: **A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other.** *Am J Hum Genet* 2004, **74:**765–769.
24. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR: **Fast and accurate genotype imputation in genome-wide association studies through pre-phasing.** *Nat Genet* 2012, **44:**955–959.
25. Willer CJ, Li Y, Abecasis GR: **METAL: fast and efficient meta-analysis of genomewide association scans.** *Bioinformatics* 2010, **26:**2190–2191.
26. Purcell S, Cherny SS, Sham PC: **Genetic power calculator: design of linkage and association genetic mapping studies of complex traits.** *Bioinformatics* 2003, **19:**149–150.
27. Allanore Y, Saad M, Dieude P, Avouac J, Distler JH, Amouyel P, Matucci-Cerinic M, Riemekasten G, Airo P, Melchers I, Hachulla E, Cusi D, Wichmann HE, Wipff J, Lambert JC, Hunzelmann N, Tiev K, Caramaschi P, Diot E, Kowal-Bielecka O, Valentini G, Mouthon L, Czirjak L, Damjanov N, Salvi E, Conti C, Muller M, Muller-Ladner U, Riccieri V, Ruiz B, *et al*: **Genome-wide scan identifies TNIP1, PSORS1C1, and RHOB as novel risk loci for systemic sclerosis.** *PLoS Genet* 2011, **7:**e1002091.
28. Beretta L, Rueda B, Marchini M, Santaniello A, Simeon CP, Fonollosa V, Caronni M, Rios-Fernandez R, Carreira P, Rodriguez-Rodriguez L, Moreno A, Lopez-Nevot MA, Escalera A, Gonzalez-Escribano MF, Martin J, Scorza R: **Analysis of class II human leucocyte antigens in Italian and Spanish systemic sclerosis.** *Rheumatology (Oxford)* 2012, **51:**52–59.
29. Harley JB, Alarcon-Riquelme ME, Criswell LA, Jacob CO, Kimberly RP, Moser KL, Tsao BP, Vyse TJ, Langefeld CD, Nath SK, Guthridge JM, Cobb BL, Mirel DB, Marion MC, Williams AH, Divers J, Wang W, Frank SG, Namjou B, Gabriel SB, Lee AT, Gregersen PK, Behrens TW, Taylor KE, Fernando M, Zidovetzki R, Gaffney PM, Edberg JC, Rioux JD, Ojwang JO, *et al*: **Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in ITGAM, PXK, KIAA1542 and other loci.** *Nat Genet* 2008, **40:**204–210.
30. Ng PC, Henikoff S: **SIFT: Predicting amino acid changes that affect protein function.** *Nucleic Acids Res* 2003, **31:**3812–3814.
31. Chun S, Fay JC: **Identification of deleterious mutations within three human genomes.** *Genome Res* 2009, **19:**1553–1561.
32. Gateva V, Sandling JK, Hom G, Taylor KE, Chung SA, Sun X, Ortmann W, Kosoy R, Ferreira RC, Nordmark G, Gunnarsson I, Svenungsson E, Padyukov L, Sturfelt G, Jonsen A, Bengtsson AA, Rantapaa-Dahlqvist S, Baechler EC, Brown EE, Alarcon GS, Edberg JC, Ramsey-Goldman R, McGwin G Jr, Reveille JD, Vila LM, Kimberly RP, Manzi S, Petri MA, Lee A, Gregersen PK: **A large-scale replication study identifies TNIP1, PRDM1, JAZF1, UHRF1BP1 and IL10 as risk loci for systemic lupus erythematosus.** *Nat Genet* 2009, **41:**1228–1233.
33. Al-Mayouf SM, Sunker A, Abdwani R, Abrawi SA, Almurshedi F, Alhashmi N, Al Sonbul A, Sewairi W, Qari A, Abdallah E, Al-Owain M, Al Motywee S, Al-Rayes H, Hashem M, Khalak H, Al-Jebali L, Alkuraya FS: **Loss-of-function variant in DNASE1L3 causes a familial form of systemic lupus erythematosus.** *Nat Genet* 2011, **43:**1186–1188.
34. Mayes MD, Bossini-Castillo L, Gorlova O, Martin JE, Zhou X, Chen WV, Assassi S, Ying J, Tan FK, Arnett FC, Reveille JD, Guerra S, Teruel M, Carmona FD, Gregersen PK, Lee AT, Lopez-Isac E, Ochoa E, Carreira P, Simeon CP, Castellvi I, Gonzalez-Gay MA, Spanish Scleroderma G, Zhernakova A, Padyukov L, Alarcon-Riquelme M, Wijmenga C, Brown M, Beretta L, Riemekasten G, *et al*: **Immunochip analysis identifies multiple susceptibility loci for systemic sclerosis.** *Am J Hum Genet* 2014, **94:**47–61.
35. Ueki M, Takeshita H, Fujihara J, Iida R, Yuasa I, Kato H, Panduro A, Nakajima T, Kominato Y, Yasuda T: **Caucasian-specific allele in non-synonymous single nucleotide polymorphisms of the gene encoding deoxyribonuclease I-like 3, potentially relevant to autoimmunity, produces an inactive enzyme.** *Clin Chim Acta* 2009, **407:**20–24.
36. Gruschwitz MS, Hornstein OP, von Den DP: **Correlation of soluble adhesion molecules in the peripheral blood of scleroderma patients with their in situ expression and with disease activity.** *Arthritis Rheum* 1995, **38:**184–189.
37. Pendergrass SA, Hayes E, Farina G, Lemaire R, Farber HW, Whitfield ML, Lafyatis R: **Limited systemic sclerosis patients with pulmonary arterial hypertension show biomarkers of inflammation and vascular injury.** *PLoS One* 2010, **5:**e12106.

doi:10.1186/s13075-014-0438-8
Cite this article as: Zochling *et al.*: An Immunochip-based interrogation of scleroderma susceptibility variants identifies a novel association at *DNASE1L3*. *Arthritis Research & Therapy* 2014 **16**:438.

Wu FT, Callisaya M, Laslett LL, Wills K, Zhou Y, et al. (2016) Lower limb muscle strength is associated with poor balance in middle-aged women: linear and nonlinear analyses. Osteoporos Int.

Pan G, Simpson S Jr, van der Mei I, Charlesworth JC, Lucas R, Ponsonby AL, Zhou Y et al. Role of genetic susceptibility variants in predicting clinical course in multiple sclerosis: a cohort study. J Neurol Neurosurg Psychiatry. 2016 Aug 24

George MF, Briggs FB, Shao X, Gianfrancesco MA, Kockum I, Harbo HF, Celius EG, Bos SD, Hedström A, Shen L, Bernstein A, Alfredsson L, Hillert J, Olsson T, Patsopoulos NA, De Jager PL, Oturai AB, Søndergaard HB, Sellebjerg F, Sorensen PS, Gomez R, Caillier SJ, Cree BA, Oksenberg JR, Hauser SL, D'Alfonso S, Leone MA, Martinelli Boneschi F, Sorosina M, van der Mei I, Taylor BV, Zhou Y et al. Multiple sclerosis risk loci and disease severity in 7,125 individuals from 10 studies. Neurol Genet. 2016 Aug 4;2(4):e87

Multiple sclerosis risk loci and disease severity in 7,125 individuals from 10 studies

OPEN

Michaela F. George, PhD
Farren B.S. Briggs, PhD
Xiaorong Shao, MA
Milena A. Gianfrancesco, MPH
Ingrid Kockum, PhD
Hanne F. Harbo, MD, PhD
Elisabeth G. Celius, MD, PhD
Steffan D. Bos, MSc, PhD
Anna Hedström, PhD
Ling Shen, PhD
Allan Bernstein, MD
Lars Alfredsson, PhD
Jan Hillert, MD, PhD
Tomas Olsson, MD, PhD
Nikolaos A. Patsopoulos, MD, PhD
Philip L. De Jager, MD, PhD
Annette B. Oturai, MD, PhD
Helle B. Søndergaard, MSc, PhD
Finn Sellebjerg, MD, PhD, DMSc
Per S. Sorensen, MD, PhD, DMSc
Refujia Gomez, BS
Stacy J. Caillier, PhD
Bruce A.C. Cree, MD, PhD, MAS
Jorge R. Oksenberg, PhD
Stephen L. Hauser, MD
Sandra D'Alfonso, PhD
Maurizio A. Leone, MD
Filippo Martinelli Boneschi, MD, PhD, MS
Melissa Sorosina, PhD

Author list continued on next page

ABSTRACT

Objective: We investigated the association between 52 risk variants identified through genome-wide association studies and disease severity in multiple sclerosis (MS).

Methods: Ten unique MS case data sets were analyzed. The Multiple Sclerosis Severity Score (MSSS) was calculated using the Expanded Disability Status Scale at study entry and disease duration. MSSS was considered as a continuous variable and as 2 dichotomous variables (median and extreme ends; MSSS of ≤5 vs >5 and MSSS of <2.5 vs ≥7.5, respectively). Single nucleotide polymorphisms (SNPs) were examined individually and as both combined weighted genetic risk score (wGRS) and unweighted genetic risk score (GRS) for association with disease severity. Random-effects meta-analyses were conducted and adjusted for cohort, sex, age at onset, and *HLA-DRB1*15:01*.

Results: A total of 7,125 MS cases were analyzed. The wGRS and GRS were not strongly associated with disease severity after accounting for cohort, sex, age at onset, and *HLA-DRB1*15:01*. After restricting analyses to cases with disease duration ≥10 years, associations were null (p value ≥0.05). No SNP was associated with disease severity after adjusting for multiple testing.

Conclusions: The largest meta-analysis of established MS genetic risk variants and disease severity, to date, was performed. Results suggest that the investigated MS genetic risk variants are not associated with MSSS, even after controlling for potential confounders. Further research in large cohorts is needed to identify genetic determinants of disease severity using sensitive clinical and MRI measures, which are critical to understanding disease mechanisms and guiding development of effective treatments. ***Neurol Genet* 2016;2:e87; doi: 10.1212/NXG.0000000000000087**

GLOSSARY

CI = confidence interval; **EAE** = experimental autoimmune encephalomyelitis; **EDSS** = Expanded Disability Severity Scale; **GRS** = unweighted genetic risk score; **GWAS** = genome-wide association studies; **KPNC** = Kaiser Permanente Medical Care Plan in the Northern California Region; **MHC** = major histocompatibility complex; **MS** = multiple sclerosis; **MSSS** = Multiple Sclerosis Severity Score; **OR** = odds ratio; **SNP** = single nucleotide polymorphism; **UCSF** = University of California at San Francisco; **wGRS** = weighted genetic risk score.

Multiple sclerosis (MS) is a severe autoimmune inflammatory disease of the central nervous system. Neurologic damage in MS is caused by irreversible demyelination of axons and lesion formation. Although early disease may manifest as attacks with full recovery, over time MS is extremely debilitating for the majority of patients. Only 37% of individuals with mild MS are employed, and within 15 years of diagnosis, 50%–60% of patients will require assistance with walking, posing tremendous economic and societal burden.[1,2]

Evidence suggests that both genetic and environmental components contribute to the risk of MS. The strongest genetic risk factor is within the human leukocyte antigen (*HLA*)-*DRB1* locus, specifically the *15:01* allele,[3] and studies support the presence of additional independent susceptibility alleles within the major histocompatibility complex (MHC) Class I and Class II regions.[4] However, these genes have not been convincingly associated with progression.[3,5]

Author affiliations are listed at the end of the article.

Funding information and disclosures are provided at the end of the article. Go to Neurology.org/ng for full disclosure forms. The Article Processing Charge was paid by the authors.

Ingrid van der Mei, PhD
Bruce V. Taylor, MD
Yuan Zhou, MS
Catherine Schaefer, PhD
Lisa F. Barcellos, PhD, MPH

Correspondence to Dr. Barcellos: lbarcellos@berkeley.edu

See editorial

Supplemental data at Neurology.org/ng

Environmental risk factors for MS include infection with Epstein-Barr virus, tobacco smoke exposure, low serum levels of vitamin D, and childhood/adolescent obesity.[6,7] With the exception of tobacco smoke and low vitamin D,[8,9] environmental influences on MS disease severity are unknown.

Through international collaboration, genome-wide association studies (GWAS) followed by replication have identified a large number of non-MHC MS risk variants.[10,11] We hypothesized that MS risk variants might also influence disease severity. We investigated the association of the Multiple Sclerosis Severity Score (MSSS) with both a weighted genetic risk score (wGRS) and unweighted genetic risk score (GRS) comprising 52 established risk variant alleles, and each susceptibility variant alone in 7,125 MS cases from 10 independent cohorts.

METHODS Study populations. Ten independent and well-characterized MS case data sets were analyzed (table 1). The analysis included 1,079 white non-Hispanic MS patients recruited from Kaiser Permanente Medical Care Plan in the Northern California Region (KPNC).[12] The following additional cases were included: 1,019 MS white non-Hispanic patients recruited from 2 other clinical sites in the United States (US1 and US2)[5]; 422 MS patients recruited through a population-based study in Oslo, Norway (Norway)[13]; 2,348 MS patients recruited through a population-based study in Sweden (Sweden)[14]; 890 MS patients from a cohort in Denmark (Denmark)[11]; 485 white non-Hispanic MS patients from a University of California San Francisco (UCSF)[15] cohort; 678 MS patients from 2 cohorts recruited in Italy (Italy1 and Italy2)[10,16]; and 204 patients from a Tasmanian cohort study (Australia).[17,18] Each case included in the meta-analysis fulfilled disease criteria for MS.[19] Classical *HLA-DRB1*15:01* typing was used in the KPNC, Sweden, and UCSF studies. Validated tagging single nucleotide polymorphisms (SNPs) for *DRB1*15:01* were used in the US1 and US2, Italy1 and Italy2, and Australia (rs9271366), Norway (rs9270986), and Denmark (rs3135388) studies. Each of these tagging SNPs was tested against the classical high-resolution *HLA-DRB1* typing performed in the KPNC data set; the correlations (r^2) were 0.99, 0.91, and 0.95, respectively.

Standard protocol approvals, registrations, and patient consents. Each study protocol was approved by the appropriate Institutional Review Board of the participating academic institution. All participants provided written informed consent.

Table 1 Characteristics of non-Hispanic white MS cases in 10 cohorts

	Mean (SD)/frequency (%)								
	KPNC	US	Norway	Sweden	Denmark	UCSF	Italy	Australia	Meta-analysis
No. of cases	1,079	1,019	422	2,348	890	485	678	204	7,125
MSSS continuous	3.37 (2.6)	3.28 (2.6)	4.59 (3.0)	4.61 (2.8)	4.48 (2.6)	2.98 (2.4)	4.57 (2.9)	4.12 (2.7)	4.09 (2.8)
MSSS binary (high)	255 (23.6)	261 (25.6)	180 (42.7)	1,045 (44.5)	370 (41.6)	101 (20.8)	295 (43.5)	73 (35.8)	2,580 (36.2)
MSSS extreme (high)	129 (19.6)	103 (17.3)	101 (40.2)	474 (41.2)	136 (36.8)	31 (10.5)	150 (40.7)	30 (28.6)	1,154 (30.4)
Female	868 (80.4)	779 (76.5)	309 (73.2)	1,702 (72.5)	614 (69.0)	333 (68.7)	436 (64.3)	147 (72.1)	5,188 (72.8)
Age at onset	31.94 (9.8)	31.03 (8.5)	32.66 (9.4)	34.60 (10.7)	30.92 (8.9)	33.47 (9.3)	32.50 (10.5)	35.25 (10.2)	32.85 (10.0)
*HLA-DRB1*15:01* (positive)	581 (53.9)	567 (55.6)	243 (57.6)	1,366 (58.2)	534 (60.0)	224 (46.2)	189 (27.9)	120 (58.8)	3,824 (53.7)
Disease duration	12.23 (8.5)	12.05 (8.5)	16.16 (10.5)	9.71 (8.6)	12.23 (7.8)	9.54 (9.1)	11.38 (8.5)	15.67 (10.2)	11.44 (8.9)
wGRS	6.75 (0.5)	6.59 (0.5)	6.84 (0.5)	6.74 (0.5)	6.68 (0.5)	6.71 (0.5)	6.85 (0.5)	6.72 (0.5)	6.73 (0.5)
GRS	56.72 (4.4)	55.77 (4.4)	57.11 (4.4)	56.44 (4.5)	56.38 (4.5)	56.46 (4.3)	57.81 (4.5)	56.47 (4.6)	56.55 (4.5)
Genotyping platform	Affymetrix GeneChip Human Mapping 500K Array set	Affymetrix GeneChip Human Mapping 500K Array set	TaqMan OpenArray Genotyping Technology	Immunochip—Illumina Infinium HD Custom Array and Illumina Human Quad 660	Immunochip—Illumina Infinium HD Custom Array	Illumina HumanHap550 Beadchip circa 2006	Illumina Human Quad 660	Illumina Infinium Hap370CNV array	—

Abbreviations: GRS = unweighted genetic risk score; KPNC = Kaiser Permanente Medical Care Plan in the Northern California Region; MS = multiple sclerosis; MSSS = Multiple Sclerosis Severity Score; SNP = single nucleotide polymorphism; UCSF = University of California at San Francisco; wGRS = weighted genetic risk score.
MSSS is presented as a continuous variable and as 2 dichotomous variables. The first dichotomous variable was based on the median MSSS value, defined as MSSS ≤5 vs >5, with a smaller score indicating a more benign phenotype. The second dichotomous variable was based on extreme ends of the MSSS distribution, defined as MSSS <2.5 vs ≥7.5. *HLA-DRB1*15:01* tag SNP if classical HLA typing was not available: US and Italy cohorts used rs9271366 as a tag SNP, Norway used rs9270986 as a tag SNP, and Denmark used rs3135388 as a tag SNP. All cohorts had data for all 52 individual SNPs, except the US and Australian cohorts. The US cohort used the following SNPs as tagging SNPs: rs6685440 tags rs11581062; and rs8106574 tags rs1077667. The Australian cohort imputed the following SNPs: rs1323292, rs7522462, rs17174870, rs10201872, rs669607, rs12212193, rs17066096, rs13192841, rs354033, rs1520333, rs10466829, rs2119704, rs7200786, rs13333054, and rs2425752. The US cohort contains 2 cohorts—US1 and US2. The Italy cohort contains 2 cohorts—Italy1 and Italy2. They are combined in table 1, but analyzed separately in the meta-analysis.

Genotyping and imputation. In addition to *HLA-DRB1*15:01* (described above), a group of 52 independent non-MHC GWAS SNPs were chosen for the present study. They were selected because they were previously identified and further replicated through GWAS, and they demonstrated larger magnitudes of effect on disease risk. All participants were genotyped using separate platforms: Affymetrix platform using the GeneChip Human Mapping 500K Array set (KPNC and United States), TaqMan OpenArray Genotyping Technology (Norway), Illumina Infinium HD Custom Array and Illumina Human Quad 660 (Sweden, Denmark, and Italy), and Illumina HumanHap550 Beadchip ©2006 (UCSF). All cohorts, except United States and Australia, contained genotyping information for all 52 variants. In the United States, tagging SNPs were used as proxies for 2 missing candidate SNPs: rs6693456 tagged rs11581062 (r^2 = 0.95), and rs8106574 tagged rs1077667 (r^2 = 0.63), based on publicly available data. In the Australia data set, 15 SNPs were missing: rs1323292, rs7522462, rs17174870, rs10201872, rs669607, rs12212193, rs17066096, rs13192841, rs354033, rs1520333, rs10466829, rs2119704, rs7200786, rs13333054, and rs2425752. Genetic data within 1-MB regions around all missing SNPs were available for each patient; therefore, imputation was possible against the 1000 Genomes reference. After imputation, missingness was lower than 2% in all cohorts. However, to recover all missing genotypes, the risk allele frequency from each cohort was used to estimate the missing genotypes. The estimation for each individual was made using a multinomial distribution generated from 1,000 random samples from each respective cohort. The probability for each of the 3 possible genotypes was generated. When a single genotype was missing for an individual, one sample was drawn randomly from the generated distribution from that cohort. This was done for a total of 146 individuals (KPNC = 92, Norway = 19, Sweden = 22, Denmark = 9, UCSF = 15, and Italy = 8) and a total of 48 SNPs (KPNC = 44, Norway = 15, Sweden = 4, Denmark = 2, UCSF = 10, and Italy = 8).

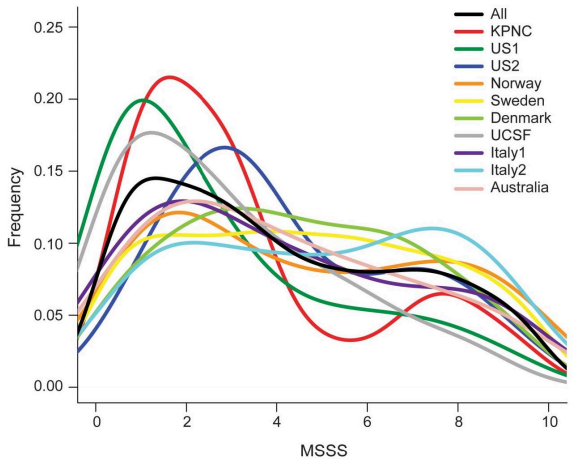
Weighted and unweighted genetic risk scores. wGRS was calculated for each MS patient using the discovery odds ratios (ORs) as the weight for each of 52 non-MHC risk alleles derived from 9,772 MS cases and 17,376 controls,[10] as previously described. In brief, the number of risk alleles for each SNP was multiplied by the weight for that variant, and then the sum across all 52 variants was calculated (table 1).[10,20] GRS was calculated as the sum of risk allele copies for each SNP without weighting (table 1). Both wGRS and GRS were analyzed as continuous variables.

Multiple Sclerosis Severity Score. MSSS is a probabilistic algorithm that uses the Expanded Disability Severity Scale (EDSS) to calculate disease severity and duration of disease, which was defined as time between first symptom and EDSS assessment.[21] The MSSS for each patient in the current study was assigned using Global MSSS reference data derived from a large independent cohort of 9,892 patients with EDSS and disease duration ranging from 1 to >30 years. EDSS scores from a pooled reference patient cohort were ranked as previously described,[21] and the average of the lowest and highest ranks for each possible EDSS value was calculated, taking into account also, scores reported for 2 years before and after each designated time point. These averages were normalized to account for the number of available assessments for each year (disease duration) and multiplied by 10. MSSS was analyzed as a continuous variable and as 2 dichotomous variables, as previously described.[5] In brief, a binary MSSS variable was based on the median MSSS value, defined as MSSS ≤5 vs >5, with a smaller score indicating more benign or "mild" disease. The second variable was based on extreme ends of the MSSS distribution, defined as MSSS <2.5 (benign) vs ≥7.5 (severe) (table 1, figure 1).

Statistical analysis. All 10 data sets were included in a random-effects meta-analysis. Random-effects meta-analysis allows for heterogeneity across studies because of inherent differences and/or differential biases among each cohort, unlike fixed-effects models, which assume a single common effect underlies each study. A random-effects meta-analysis is generally more conservative, generating wider confidence intervals (CIs) and larger p values. Weighted and unweighted GRS, and all 52 non-MHC risk variants, were tested with the 3 MSSS outcomes in the meta-analysis.

In addition, analyses restricted to cases with a preestablished disease duration greater than or equal to 10 years were conducted to increase the likelihood that the MSSS measurement was stable. Both adjusted linear and logistic regression models were used to estimate adjusted beta values (β), ORs, and 95% CIs. The meta-analysis was adjusted for sex, age at onset, and *HLA-DRB1*15:01*. All analyses were conducted in STATA v13.1 (StataCorp, College Station, TX). The present study had sufficient power for all analyses (table e-1 at Neurology.org/ng).

Figure 1 Density plot of MSSS by cohort



The density plot represents the frequency distribution of MSSS in each of the 10 cohorts and the total 7,125 individuals with MS. KPNC = Kaiser Permanente Medical Care Plan in the Northern California Region; MS = multiple sclerosis; MSSS = Multiple Sclerosis Severity Score. UCSF = University of California at San Francisco.

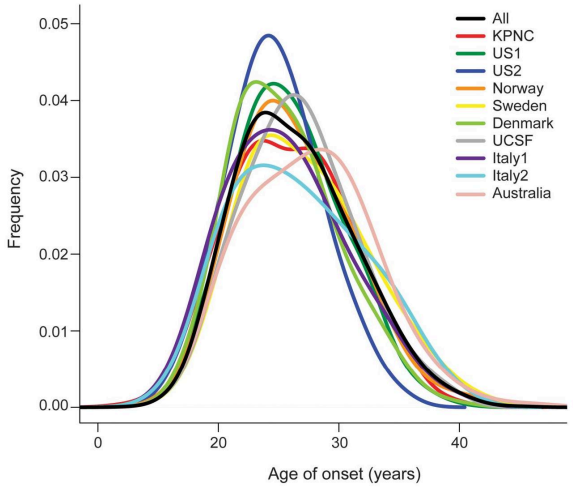
RESULTS A total of 7,125 individuals were included in the meta-analysis of the 3 MSSS outcomes (table 1). The overall distributions of MSSS demonstrated similarities across individual cohorts (figure 1) and were comparable to those in other cohorts reported previously[21]; however, there were some notable

differences. Cohorts from Nordic countries and Italy had a larger proportion of cases with greater disease severity; disease duration, on average, was shorter in the Swedish and US (UCSF) cohorts (table 1). The sex distribution was also very similar across cohorts, with a 3:1 female to male ratio; KPNC had the highest proportion of females (80.4%), and Italy had the lowest proportion of females (64.3%) (table 1). Age at onset was similar among all cohorts and normally distributed, ranging from 30.9 years of age in Denmark to 35.3 years of age in Australia (table 1, figure 2). On average, disease duration since first symptom was 11.4 years across all cohorts and normally distributed, ranging from 9.5 years in UCSF to 16.2 years in Norway (table 1, figure e-1). Last, the distribution of *HLA-DRB1*15:01* was typical of established genetic patterns in the literature (table 1): MS participants of Northern European descent were more likely to be *HLA-DRB1*15:01* positive than participants in the Italy cohort.[22] The non-MHC wGRS and GRS were similar across cohorts (figures e-2 and e-3).

Meta-analyses accounting for the random effect of each cohort were used. We observed some evidence for association with wGRS and GRS for the MSSS outcomes (table 2). However, after restricting the data set to individuals who had a disease duration of 10 years or more, no significant associations remained (table e-2). Sex, age at onset, and *HLA-DRB1*15:01* status were all analyzed as fixed effects. Sex and age at onset were consistently associated with MSSS in all 3 models, even after restricting to individuals with 10 years or more of disease (all p values <0.001) (table 2 and table e-2). Male sex and a later age at onset were both associated with more severe disease. *HLA-DRB1*15:01* was not associated with MSSS in any of the models.

The 52 non-MHC risk variants were also tested individually in random-effects meta-analytic models adjusting for sex, age at onset, and *HLA-DRB1*15:01* across all MS cases (full cohort), and in MS cases with disease duration of 10 years or more (restricted cohort) (table 3, table e-3). In the full cohort, 2 variants showed evidence of association for all 3 MSSS outcomes: rs874628 and rs650258, within *MPV17L2* and 44-kb upstream of *CD6*, respectively. Similar results were observed in the restricted cohort (data not shown). All associations were in the same direction as the wGRS and GRS. However, after accounting for multiple testing, no single variant remained significant. When analyses were stratified by *DRB1*15:01*-positive and -negative carrier status in cases, results were similar (data not shown). We also restricted the above analyses to combined US, European, and Australian cohorts separately and did not observe any cohort-specific associations between MSSS and the wGRS or GRS (data not shown).

Figure 2 Density plot of age at onset by cohort



The density plot represents the frequency distribution of age at onset in each of the 10 cohorts and the total 7,125 individuals with MS. KPNC = Kaiser Permanente Medical Care Plan in the Northern California Region; MS = multiple sclerosis; MSSS = Multiple Sclerosis Severity Score. UCSF = University of California at San Francisco.

DISCUSSION The identification of disease-modifying genes in MS is critical to further our understanding of disease pathogenesis, given the phenotype heterogeneity observed in patients for clinical manifestations, temporal course, and disease severity. This important topic has been previously reviewed, and several challenges have been described.[23] Minimal progress has been made in the past decade owing mostly to the need for large, well-designed studies to detect presumably modest genetic effects, similar to what has been observed for GWAS, and clinical outcome data for large numbers of patients that accurately capture measures likely to be modified by genetic variation. Here, a meta-analysis of association between 52 established non-MHC MS genetic risk factors with larger effect sizes demonstrated in GWAS and disease severity was performed. Thus far, this is the largest combined MS patient cohort used to evaluate the relationship between established risk variants and disease severity. Ten independent and well-characterized data sets comprising 7,125 individuals with clinically definite MS were studied. The hypothesis that one or more known MS genetic risk variants are also associated with disease severity characterized by MSSS was comprehensively tested. Genetic factors that

Table 2 Meta-analysis results for wGRS and GRS and MSSS in 7,125 MS cases

	MSSS outcome					
	Continuous		Binary		Extreme[a]	
	β (95% CI)	*p* Value	OR (95% CI)	*p* Value	OR (95% CI)	*p* Value
wGRS (unadjusted)	0.08 (−0.1 to 0.2)	0.22	1.1 (1.0–1.2)	0.36	1.1 (1.0–1.3)	0.17
wGRS (adjusted)	0.1 (0.01 to 0.3)	0.03	1.1 (1.0–1.2)	0.09	1.2 (1.0–1.4)	0.02
Sex	−0.6 (−0.7 to −0.4)	<0.001	0.7 (0.6–0.8)	<0.001	0.5 (0.5–0.6)	<0.001
Age at onset	0.06 (0.05 to 0.07)	<0.001	1.0 (1.0–1.0)	<0.001	1.1 (1.1–1.1)	<0.001
*HLA-DRB1*15:01*	−0.02 (−0.1 to 0.1)	0.76	1.0 (0.9–1.1)	0.65	1.0 (0.8–1.1)	0.75
GRS (unadjusted)	0.01 (−0.004 to 0.02)	0.18	1.0 (1.0–1.0)	0.32	1.0 (1.0–1.0)	0.11
GRS (adjusted)	0.02 (0.002 to 0.03)	0.03	1.0 (1.0–1.0)	0.09	1.0 (1.0–1.0)	0.01
Sex	−0.6 (−0.7 to −0.4)	<0.001	0.7 (0.6–0.8)	<0.001	0.5 (0.5–0.6)	<0.001
Age at onset	0.06 (0.05 to 0.07)	<0.001	1.0 (1.0–1.0)	<0.001	1.1 (1.1–1.1)	<0.001
*HLA-DRB1*15:01*	−0.02 (−0.1 to 0.1)	0.77	1.0 (0.9–1.1)	0.66	1.0 (0.8–1.1)	0.76

Abbreviations: CI = confidence interval; GRS = unweighted genetic risk score; MS = multiple sclerosis; MSSS = Multiple Sclerosis Severity Score; OR = odds ratio; wGRS = weighted genetic risk score.
All βs and ORs are from the random-effects models. Unadjusted and adjusted results are shown. Adjusted models included the following covariates: sex, age at onset, and *HLA-DRB1*15:01*. All βs and ORs for the covariates in the adjusted models are also listed above.
[a] A total of 3,795 individuals with MS are included in this analysis.

influence disease susceptibility tested here were not shown convincingly to affect disease severity.

The individuals included in this analysis are representative of the international MS population with regard to sex distribution, average age at onset, and proportion of *HLA-DRB1*15:01* allele carriers. Male sex and older age at onset were associated with more severe disease, which is consistent with other reports.[24] While *HLA-DRB1*15:01* is an established risk factor for MS, it was not associated with MS disease severity here, as in previous studies that examined MSSS[3,5] and brain atrophy.[25]

Available MS family data show that familial factors do not substantially influence eventual disease severity. However, they may increase the probability of a progressive clinical course, either from onset or after a phase of relapsing remitting disease, and evidence suggests that genetic influences are involved.[26] Findings from studies of other neurodegenerative diseases, such as Alzheimer disease and Parkinson disease, have linked disease progression with specific genetic markers, also raising the possibility that MS progression might have a genetic component,[27,28] given evidence for both neurodegeneration and inflammation in MS. However, to date, there has been limited success identifying non-MHC genetic associations with disease severity or other clinical phenotypes in MS. One study showed that a small subset (n = 16) of risk variants investigated in the present study was not associated with clinical and MRI outcomes in 179 recently diagnosed MS patients.[29] However, the cumulative burden of non-MHC risk variants (n = 110) contributed modestly to the number of relapses in 842 MS patients.[30] Furthermore, some variants studied (n = 17) were associated with attack severity, recovery, or frequency in 503 recently diagnosed MS cases.[31] Relapse data were not available for the present study. These observations require further investigation and much larger samples sizes for confirmation. The most compelling evidence for genetic associations with cerebrospinal fluid antibody levels, both immunoglobulin G index and oligoclonal band positive status, was reported recently for MHC and immunoglobulin heavy-chain region genes in 6,950 patients.[32] Both antibodies are markers of more severe disease in MS. The results, in conjunction with the present study findings, underscore the need for new studies to identify or exclude genetic contributions to disease severity in MS.

There is evidence in animal models related to MS to suggest that risk alleles have an effect on progression. The experimental autoimmune encephalomyelitis (EAE) model, which mirrors an inflammatory autoimmune disease of the central nervous system in rats and mice, has offered numerous experimental insights into MS. When genetically dissected into high-resolution quantitative trait loci, *Eae25* and *Eae29* have been shown to influence both susceptibility and progression.[33,34] In addition, differential expression of an interleukin 2 (*IL2*) repressor in the gene *ZEB1* results in EAE severity changes.[35] Similarly, congenic rats with *Eae18b* locus have been

Table 3 Meta-analysis results for 52 SNPs and MSSS in 7,125 MS cases

			MSSS outcome					
			Continuous		Binary		Extreme[a]	
Chr	Gene	SNP	β (95% CI)[b]	p Value	OR (95% CI)[b]	p Value	OR (95% CI)[b]	p Value
1	*MMEL1*	rs4648356	−0.07 (−0.2 to 0.02)	0.14	1.0 (0.9–1.1)	0.45	0.9 (0.8–1.0)	0.14
	EVI5	rs11810217	0.06 (−0.03 to 0.2)	0.21	1.1 (1.0–1.1)	0.27	1.1 (1.0–1.2)	0.15
	VCAM1	rs11581062	0.05 (−0.05 to 0.1)	0.32	1.0 (1.0–1.1)	0.24	1.0 (0.9–1.2)	0.65
	CD58	rs1335532	0.07 (−0.07 to 0.2)	0.32	1.1 (0.9–1.2)	0.44	1.1 (0.9–1.3)	0.35
	RGS1	rs1323292	−0.1 (−0.2 to 0.01)	0.06	0.9 (0.8–1.0)	0.01	1.0 (0.8–1.1)	0.48
	C1orf106	rs7522462	−0.09 (−0.2 to 0.01)	0.09	0.9 (0.9–1.0)	0.08	0.9 (0.8–1.0)	0.08
2	No gene	rs12466022	0.05 (−0.05 to 0.2)	0.34	1.0 (1.0–1.1)	0.49	1.0 (0.9–1.2)	0.54
	PLEK	rs7595037	0.05 (−0.03 to 0.1)	0.23	1.0 (1.0–1.1)	0.55	1.1 (1.0–1.2)	0.16
	MERTK	rs17174870	−0.003 (−0.1 to 0.10)	0.94	1.0 (0.9–1.1)	0.71	1.0 (0.9–1.2)	0.80
3	*SP140*	rs10201872	0.01 (−0.1 to 0.1)	0.82	1.0 (0.9–1.1)	0.69	1.0 (0.9–1.2)	0.99
	EOMES	rs11129295	0.02 (−0.07 to 0.1)	0.64	1.0 (0.9–1.1)	0.98	1.0 (0.9–1.1)	0.93
	No gene	rs669607	0.09 (−0.003 to 0.2)	0.05	1.1 (0.9–1.1)	0.12	1.1 (1.0–1.2)	0.08
	CBLB	rs2028597	0.06 (−0.1 to 0.2)	0.49	1.0 (0.9–1.2)	0.68	1.0 (0.8–1.2)	0.88
	TMEM39A	rs2293370	−0.1 (−0.2 to 0.02)	0.09	0.9 (0.8–1.0)	0.07	0.9 (0.8–1.1)	0.18
	CD86	rs9282641	−0.05 (−0.2 to 0.1)	0.56	1.0 (0.9–1.1)	0.69	1.0 (0.8–1.2)	0.76
	IL12A	rs2243123	0.01 (−0.08 to 0.1)	0.82	1.0 (0.9–1.1)	0.69	1.0 (0.9–1.1)	0.91
5	*IL7R*	rs6897932	0.004 (−0.1 to 0.1)	0.93	1.0 (0.9–1.1)	0.61	1.0 (0.9–1.1)	0.76
	PTGER4	rs4613763	0.06 (−0.06 to 0.2)	0.31	1.1 (1.0–1.2)	0.08	1.1 (1.0–1.3)	0.15
	IL12B	rs2546890	0.05 (−0.03 to 0.1)	0.24	1.0 (0.9–1.1)	0.90	1.1 (1.0–1.2)	0.06
6	*BACH2*	rs12212193	0.03 (−0.06 to 0.1)	0.56	1.0 (0.9–1.1)	0.80	1.0 (0.9–1.2)	0.58
	THEMIS	rs802734	0.03 (−0.07 to 0.1)	0.58	1.0 (0.9–1.1)	0.67	1.1 (0.9–1.2)	0.40
	MYB	rs11154801	0.08 (−0.01 to 0.2)	0.08	1.1 (1.0–1.2)	0.10	1.1 (1.0–1.2)	0.15
	IL22RA2	rs17066096	0.09 (−0.01 to 0.2)	0.09	1.1 (1.0–1.2)	0.14	1.2 (1.0–1.3)	0.01
	No gene	rs13192841	−0.02 (−0.1 to 0.08)	0.71	1.0 (0.9–1.1)	0.77	1.0 (0.9–1.1)	0.52
	TAGAP	rs1738074	−0.03 (−0.1 to 0.06)	0.50	1.0 (0.9–1.1)	0.87	1.0 (0.9–1.1)	0.54
7	*ZNF746*	rs354033	−0.03 (−0.1 to 0.07)	0.56	1.0 (0.9–1.1)	0.86	0.9 (0.8–1.1)	0.31
8	*IL7*	rs1520333	−0.01 (−0.1 to 0.09)	0.89	1.0 (1.0–1.1)	0.30	1.0 (0.9–1.1)	0.72
	MYC	rs4410871	0.06 (−0.04 to 0.2)	0.25	1.1 (1.0–1.1)	0.28	1.0 (0.9–1.2)	0.54
	PVT1	rs2019960	0.05 (−0.05 to 0.2)	0.32	1.0 (1.0–1.1)	0.49	1.1 (0.9–1.2)	0.38
10	*IL2RA*	rs3118470	0.05 (−0.04 to 0.1)	0.30	1.0 (0.9–1.1)	0.62	1.1 (1.0–1.2)	0.07
	ZMIZ1	rs1250550	0.04 (−0.05 to 0.1)	0.37	1.1 (1.0–1.1)	0.24	1.1 (1.0–1.2)	0.14
	HHEX	rs7923837	0.02 (−0.07 to 0.1)	0.69	1.0 (1.0–1.1)	0.56	1.0 (0.9–1.1)	0.80
11	*CD6*	rs650258	0.1 (0.02 to 0.2)	0.02	1.1 (1.0–1.2)	0.02	1.2 (1.0–1.3)	0.01
12	*TNFRSF1A*	rs1800693	−0.02 (−0.1 to 0.1)	0.63	1.0 (0.9–1.1)	0.78	1.0 (0.9–1.1)	0.60
	CLECL1	rs10466829	0.02 (−0.06 to 0.1)	0.61	1.0 (0.9–1.1)	0.84	1.1 (1.0–1.2)	0.23
	CYP27B1	rs12368653	−0.04 (−0.1 to 0.05)	0.42	1.0 (0.9–1.1)	0.61	1.0 (0.9–1.1)	0.43
	ARL61P4	rs949143	0.07 (−0.03 to 0.2)	0.16	1.0 (1.0–1.1)	0.55	1.1 (1.0–1.2)	0.16
14	*ZFP36L1*	rs4902647	0.06 (−0.03 to 0.1)	0.19	1.0 (1.0–1.1)	0.29	1.1 (1.0–1.2)	0.15
	BATF	rs2300603	−0.05 (−0.2 to 0.05)	0.31	1.0 (0.9–1.0)	0.34	1.0 (0.9–1.1)	0.79
	GALC	rs2119704	−0.01 (−0.2 to 0.2)	0.93	1.0 (0.9–1.2)	0.77	0.9 (0.7–1.1)	0.44
16	*CLEC16A*	rs7200786	0.02 (−0.06 to 0.1)	0.59	1.0 (1.0–1.1)	0.67	1.1 (1.0–1.2)	0.28
	IRF8	rs13333054	0.1 (−0.01 to 0.2)	0.07	1.1 (1.0–1.2)	0.05	1.1 (1.0–1.3)	0.09

Continued

Table 3 Continued

			MSSS outcome					
			Continuous		Binary		Extreme[a]	
Chr	Gene	SNP	β (95% CI)[b]	p Value	OR (95% CI)[b]	p Value	OR (95% CI)[b]	p Value
17	*STAT3*	rs9891119	−0.02 (−0.1 to 0.07)	0.63	1.0 (0.9–1.1)	0.93	1.0 (0.9–1.1)	0.95
18	*MALT1*	rs7238078	−0.06 (−0.2 to 0.04)	0.24	1.0 (0.9–1.0)	0.27	1.0 (0.8–1.1)	0.44
19	*TNFSF14*	rs1077667	−0.04 (−0.2 to 0.07)	0.47	1.0 (0.9–1.0)	0.28	1.0 (0.9–1.2)	0.98
	TYK2	rs8112449	−0.01 (−0.1 to 0.09)	0.85	1.0 (0.9–1.1)	0.53	0.9 (0.8–1.0)	0.20
	MPV17L2	rs874628	0.1 (0.02 to 0.2)	0.02	1.1 (1.0–1.2)	0.03	1.2 (1.0–1.3)	0.01
	DKKL1	rs2303759	0.06 (−0.04 to 0.2)	0.24	1.1 (1.0–1.2)	0.05	1.0 (0.9–1.2)	0.51
20	*CD40*	rs2425752	0.03 (−0.1 to 0.06)	0.52	1.0 (0.9–1.0)	0.27	1.0 (0.9–1.1)	0.58
	CYP24A1	rs2248359	0.07 (−0.02 to 0.2)	0.13	1.1 (1.0–1.1)	0.20	1.1 (0.9–1.2)	0.39
22	*MAPK1*	rs2283792	0.06 (−0.03 to 0.1)	0.19	1.0 (1.0–1.1)	0.28	1.1 (1.0–1.2)	0.18
	SCO2	rs140522	−0.02 (−0.1 to 0.08)	0.71	1.0 (0.9–1.1)	0.99	1.1 (0.9–1.2)	0.36

Abbreviations: CI = confidence interval; MS = multiple sclerosis; MSSS = Multiple Sclerosis Severity Score; OR = odds ratio; SNP = single nucleotide polymorphism.
These marginal models tested the association between the 3 MSSS phenotypes and all 52 individual SNPs. Cohort was a random-effect variable in the models.
[a] A total of 3,795 individuals with MS are included in this analysis.
[b] All βs and ORs are from the adjusted models. Models are adjusted for sex, age at onset, *HLA-DRB1*15:01*, and cohort.

shown to develop milder disease, with decreased demyelination and reduced recruitment of inflammatory cells to the brain.[36] Evidence for multiple linked quantitative trait loci within the Tmevd2/Eae interval controlling disease severity in mice has also been reported.[37] However, these findings have not been replicated in humans.

The present study had many strengths. First, the MSSS has been favored over EDSS to capture disease severity in MS because it incorporates disease duration to account for time.[21] This approach builds on earlier work[38] and is advantageous for large studies where EDSS measurements may have only been recorded at a single time point. We used data from a large, independent cohort reference (Global MSSS) to assign MSSS for each case in the present study. Sparse data were therefore avoided for cells in the reference table and potentially much larger variances, and missing data, in general; both could result from small sample sizes if our cohorts, which vary greatly in size, had been characterized instead by local MSSS.[21] Furthermore, the large sample size used to create the Global MSSS made it possible to reduce the effect of stochastic fluctuations over time. Specifically, EDSS scores in the surrounding 2 years for each disease duration year are accounted for in the ranking; very large data sets are required to do this. It is important to note that the Global MSSS reference data set is at least ~5 times larger than the largest individual cohort in the present study. The power of Global MSSS for genetic studies based on the large pooled reference data has also been previously demonstrated.[21]

Additional strengths included application of a statistical model for association testing that accounted for random (cohort) effects, and we adjusted for potential confounders such as age and sex. Moreover, use of MSSS extremes analysis reduced the possibility of phenotype misclassification because individuals with benign disease were compared with individuals with severe disease, and individuals in the middle of the MSSS spectrum who are most likely to be misclassified were excluded. While restricting analyses to extreme categories of MSSS reduced the overall number of individuals in our data set, given our overall large sample size, close to 4,000 individuals were still available for association testing.

Analyses were also restricted to individuals with 10 or more years of disease duration since symptom onset, to help ensure stability of the MSSS. At least 1 year of duration is required before the EDSS can be used to calculate the MSSS with the published algorithm.[21] Median time to requiring unilateral assistance ranges from 15 to 30 years in MS, based on the present estimates.[24] Conservatively, the present study used disease duration near the median value for the overall data set (table 1), but before the early end of the transition to requiring unilateral assistance (or an EDSS of 6). When analyses were restricted by disease duration (≥5, ≥10, ≥15, and ≥20 years), wGRS, GRS, and individuals SNPs (data not shown) were

not associated with any of the 3 MSSS outcomes (figures e-4 and e-5).

There were also several limitations in this study. The EDSS used to calculate the MSSS is heavily weighted by physical measurements of disability. While MSSS is a widely accepted measure of disability for MS research, there are concerns about its sensitivity, given that different combinations of disease duration and EDSS can result in similar measures of the MSSS. In addition, it is possible that population-specific EDSS rankings could differ from those derived using the Global MSSS. A genetic association might therefore only be uncovered when a population is examined independently. Even larger studies will be required to determine whether studies of disability based on locally derived values for MSSS reveal population-specific genetic effects. Furthermore, because subclinical disease can occur many years before clinical symptom onset, the measure of disease duration required to derive the MSSS may not be accurate. Recent work shows that age could potentially be used for ranking EDSS in future studies when disease duration is unknown, because the exact age of symptom onset may not be certain.[39] Longitudinal data on progression were not available for cases in the present study. There is also potential selection bias if individuals with mild disease are more willing to participate in research studies than those with severe disease. However, more than 1,100 participants were categorized in the highest or "most severe" MSSS phenotype in the present study (table 1), which provided a sufficient representation of cases with severe disease. Finally, our negative findings may have resulted from population or individual factors that could be not assessed in the present study. Additional genetic variants (untested), environmental exposures, or medication histories were not available and may distinguish MSSS differences in the presence or absence of the established genetic risk variants that were investigated.

The current data set includes 52 non-MHC MS genetic risk variants,[10] and a larger list of the most recently identified non-MHC MS risk variants through GWAS and follow-up studies is now available.[11] The 52 non-MHC MS risk variants used here were stronger genetic markers of MS susceptibility; the average OR based on GWAS for our 52 non-MHC variants is 1.19 vs average OR for new 48 non-MHC variants, which is 1.09.[10,11] GWAS of disease severity, to date, have not yet yielded new candidates at genome-wide significance[5,10]; and while disease risk and severity variants appear to differ in MS, this still remains a question. We have comprehensively pursued testing of the strongest established disease risk variants identified, thus far, for evidence of influence on MS severity. Larger studies of cases with additional genotype data for newly discovered GWAS variants are warranted. Finally, whole-genome data were not available to formally adjust for potential effects of population stratification. However, this was not a case-control study, each cohort was treated separately in the analysis, and consistent nonsignificant estimates for each of the 10 independent cohorts were observed. Our findings in this large, well-powered study are unlikely to result from population stratification.

Results derived from investigation of a large number of recently established GWAS variants in 7,125 MS cases suggest that the genetics underlying MS susceptibility and disease severity, as measured by MSSS, do not substantially overlap. Sensitive measures of severity and progression are needed, and for larger prospective cohort studies of incident cases. Comprehensive quantitative trait data derived from high-resolution brain MRIs can be used as dependent variables for a GWAS, as recently demonstrated with 284 MS patients,[40] or for whole-genome sequencing studies. Longitudinal studies will be informative for genetic investigations. However, such studies will be necessarily balanced by a lack of statistical power to detect modest genetic effects unless very large numbers of individuals can be assembled and followed for long periods of time.

AUTHOR AFFILIATIONS

From the Division of Epidemiology (M.F.G., X.S., M.A.G., L.F.B), Genetic Epidemiology and Genomics Laboratory, School of Public Health, University of California, Berkeley; Department of Epidemiology and Biostatistics, School of Medicine (F.B.S.B.), Case Western Reserve University, Cleveland, OH; Department of Clinical Neuroscience (I.K., A.H., J.H, T.O.), Karolinska Institutet, Stockholm, Sweden; Department of Neurology (H.F.H., E.G.C., S.D.B.), Oslo University Hospital; Institute of Clinical Medicine (H.F.H., S.D.B.), University of Oslo, Norway; Kaiser Permanente Division of Research (L.S., C.S., L.F.B.), Oakland, CA; Sonoma West Medical Center (A.B.), Sebastopol, CA; Institute of Environmental Medicine (L.A.), Karolinska Institutet, Stockholm, Sweden; Centre for Occupational and Environmental Medicine (L.A.), Stockholm County Council, Sweden; Program in Translational NeuroPsychiatric Genomics (N.A.P., P.L.D.J.), Institute for the Neurosciences, Department of Neurology, Division of Genetics (N.A.P., P.L.D.J.), Department of Medicine, Brigham & Women's Hospital, Harvard Medical School (N.A.P., P.L.D.J.), Boston, MA; Broad Institute of Harvard and Massachusetts Institute of Technology (N.A.P., P.L.D.J.), Cambridge; Danish Multiple Sclerosis Center (A.B.O., H.B.S., F.S., P.S.S.), Department of Neurology, Rigshospitalet, University of Copenhagen, Denmark; Department of Neurology (R.G., S.J.C., B.A.C.C., J.R.O., S.L.H.), University of California, San Francisco; Department of Health Sciences (S.D.A.), Interdisciplinary Research Center of Autoimmune Diseases IRCAD (S.D.A., M.A.L.), University of Eastern Piedmont, Novara, Italy; SCDU Neurology (M.A.L.), AOU "Maggiore della Carità," Novara, Italy; Laboratory of Genetics of Neurological Complex Disorders (F.M.B., M.S.), Institute of Experimental Neurology (INSPE), Division of Neuroscience, San Raffaele Scientific Institute, Milan, Italy; Department of Neurology (F.M.B.), Institute of Experimental Neurology (INSPE), Division of Neuroscience, San Raffaele Scientific Institute, Milan, Italy; and Menzies Institute for Biomedical Research (I.v.d.M., B.V.T., Y.Z.), University of Tasmania, Hobart, Australia.

AUTHOR CONTRIBUTIONS

Dr. George: conception and design, analysis and interpretation of the data, drafting of the manuscript, and final approval of the version to be published. Dr. Briggs: conception and design, analysis and interpretation of the data, reviewing and critically revising the manuscript, and final approval of the version to be published. Ms. Shao and Ms. Gianfrancesco: analysis and interpretation of the data, reviewing and critically revising the manuscript, final approval of the version to be published. Dr. Kockum and Dr. Harbo: acquisition of the data, clinical evaluation of patients, collection of samples, evaluation of manuscript, and final approval of the version to be published. Dr. Celius: clinical evaluation of patients, collection of samples, evaluation of manuscript, and final approval of the version to be published. Dr. Bos, Dr. Hedström, and Dr. Shen: data preparation, evaluation of manuscript, and final approval of the version to be published. Dr. Bernstein: clinical evaluation of patients, evaluation of manuscript, and final approval of the version to be published. Dr. Alfredsson, Dr. Hillert, Dr. Olsson, Dr. Patsopoulos, Dr. De Jager, Dr. Oturai, Dr. Søndergaard, Dr. Sellebjerg, and Dr. Sorensen: acquisition of the data, collection of samples, evaluation of manuscript, and final approval of the version to be published. Ms. Gomez, Ms. Caillier, Dr. Cree, and Dr. Hauser: clinical evaluation of patients, evaluation of manuscript, and final approval of the version to be published. Dr. Oksenberg: acquisition of the data, collection of samples, evaluation of manuscript, and final approval of the version to be published. Dr. D'Alfonso: acquisition of the data, clinical evaluation of patients, collection of samples, evaluation of manuscript, and final approval of the version to be published. Dr. Leone: acquisition of the data, collection of samples, evaluation of manuscript, and final approval of the version to be published. Dr. Martinelli Boneschi: acquisition of the data, clinical evaluation of patients, collection of samples, evaluation of manuscript, and final approval of the version to be published. Dr. Sorosina: acquisition of the data, collection of samples, evaluation of manuscript, and final approval of the version to be published. Dr. van der Mei and Dr. Taylor: acquisition of the data, clinical evaluation of patients, collection of samples, evaluation of manuscript, and final approval of the version to be published. Mr. Zhou: data preparation, evaluation of manuscript, and final approval of the version to be published. Dr. Shaefer: conception and design, acquisition of the data, reviewing and critically revising the manuscript, and final approval of the version to be published. Dr. Barcellos: conception and design, acquisition of the data, analysis and interpretation of the data, drafting and writing of the manuscript, and final approval of the version to be published.

STUDY FUNDING

This work was supported by the NIH/National Institute of Neurological Disorders and Stroke R01NS049510, NIH/National Institute of Allergy and Infectious Diseases R01AI076544, and NIH/National Institute of Environmental Health Sciences R01 ES017080; Foundation for Multiple Sclerosis (grants "Progetto Speciale Immunochip," 2011/R/14, Fondazione cariplo [2010-0728]); and NIH/National Institute of Neurological Disorders and Stroke R01NS026799 and NIH/National Institute of Neurological Disorders and Stroke R01NS049477.

DISCLOSURE

Dr. George, Dr. Briggs, and Ms. Shao report no disclosures. Ms. Gianfrancesco has received research support from NIH/National Institute of Neurological Disorders and Stroke. Dr. Kockum has received speaker honoraria from Merc Serono; has received research support from NHR Fonden and Karolinska University Hospital; and holds stock in Novo Nordisk and several Swedish companies (all stock unrelated to this study). Dr. Harbo has received research support and/or travel funding from Biogen Idec, Sanofi-Aventis, TEVA, and Novartis and has received research support from the Norwegian Research Council, Norwegian Health Authorities South-East, the University of Oslo, the Oslo Multiple Sclerosis Society, the Odd Fellow MS Society (Norway), and Novartis. Dr. Celius has received unrestricted research grants from Novartis and Biogen Idec and payment for advice and lectures from Almirall, Biogen Idec, Genzyme, Sanofi, TEVA, Merck-Serono, and Novartis. Dr. Bos, Dr. Hedström, and Dr. Shen report no disclosures. Dr. Bernstein receives research funding from Eli Lilly and Roach (for Alzheimer studies) and has received research support from Genentech, Biogen, Lundbeck, Accera, and AZ Therapeutics. Dr. Alfredsson has received speaker honoraria from TEVA and Biogen Idec and has received research support from the following government entities: the Swedish Medical Research Council; the Swedish Council for Working Life, Health and Welfare; AFA Insurance; and the Swedish Brain Foundation. Dr Hillert has received honoraria for serving on advisory boards for BiogenIdec, Sanofi-Genzyme, and Novartis; has received speaker fees from BiogenIdec, Merck-Serono, Bayer-Schering, TEVA, Novartis, and Sanofi-Genzyme; has received research support from BiogenIdec, Merck-Serono, TEVA, Novartis, Sanofi-Genzyme, Bayer-Schering, the Swedish Research Council, the Karolinska Institute, and the Swedish Brain Foundation; and has served on the editorial board of the *Journal of Neuroimmunology*. Dr. Olsson has served on the scientific advisory boards of/received speaker honoraria from Biogen Idec, Merck, Novartis, TEVA, Sanofi-Aventis, Genzyme, Almirall, and Actelion; has received research support from Biogen Idec, Bayer Schering, Genzyme, Merck, Sanofi-Aventis, Novartis, AstraZeneca, the Swedish Research Council, the Knut and Alice Wallenbergs Foundation, the AFA Foundation, and the Swedish Brain Foundation; and has served on the editorial board of *Current Opinion in Immunology*. Dr. Patsopoulos has served as a consultant for Merck and has received research support from Intel Corporation, the National Multiple Sclerosis Society, and the Harvard NeuroDiscovery Center. Dr. De Jager has served on the scientific advisory boards of TEVA Neuroscience and Sanofi/Genzyme; has received travel funding/speaker honoraria from Biogen Idec, Source Healthcare Analytics Pfizer Inc, and TEVA; has served on editorial boards of the *Journal of Neuroimmunology*, *Neuroepigenetics*, and *Multiple Sclerosis*; and has received research support from Biogen Idec, Vertex, Sanofi/Genzyme, and the National MS Society. Dr. Oturai has served on the scientific advisory boards for Novartis, Sanofi/Genzyme, and Biogen Idec; has received speaker honoraria from Biogen Idec, Merck Serono, TEVA, Sanofi/Genzyme, and Novartis; and has received research support from Biogen Idec, Novartis, the Danish Council for Strategic Research, Brdr. Rønje Holding, the Danish Multiple Sclerosis Society, the Warwara Larsen Foundation, and the Johnsen Foundation. Dr. Søndergaard has received travel funding/speaker honoraria from Biogen Idec, Genzyme, and TEVA and has received research support from Sclerose Foreningen, MS Networks and Immunogenetics, and A.P. Møller Fonden-Fonden til lægevidenskabens fremme. Dr. Sellebjerg has served on scientific advisory boards, been on the steering committees of clinical trials, served as a consultant, received support for congress participation, received speaker honoraria, or received research support for his laboratory from Bayer Schering, Biogen Idec, Genzyme, Merck Serono, Novo Nordisk, Novartis, Sanofi-Aventis, Schering Plough, TEVA, the Danish Strategic Research Council, the Danish Multiple Sclerosis Society, and the Lounkær Foundation and has served on the editorial board of *Multiple Sclerosis and Related Disorders*. Dr. Sorensen has served on scientific advisory boards for/received speaker honoraria from Biogen Idec, Merck Serono, Novartis, Genmab, TEVA, GlaxoSmithKline, Bayer Schering, Sanofi-Aventis, and Genzyme; has received research support from Biogen Idec, Bayer Schering, Merck Serono, TEVA, Sanofi-Aventis, Novartis, RoFAR, Roche, Genzyme, the Danish Multiple Sclerosis Society, the Danish Medical Research Council, and the European Union Sixth Framework Programme: Life sciences, Genomics and Biotechnology for Health; and has served on the editorial boards of the *European Journal of Neurology*, *Therapeutic Advances in Neurological Disorders*, and *Multiple Sclerosis Journal*. Ms. Gomez and Ms. Caillier report no disclosures. Dr. Cree has been a consultant for Abbvie, Biogen Idec, EMD Serono, Genzyme/Sanofi-Aventis, MedImmune, Shire, Novartis, and TEVA Neurosciences; has received research support from Acorda, Avanir, Biogen Idec, EMD Serono, Hoffman La Roche, MedImmune, Novartis, Receptos, and TEVA; and has served on the editorial board of *Annals of Neurology*. Dr. Oksenberg has served on the editorial board of *Genes and Immunity*; has been an employee of the University of California at San Francisco; and has received research support from NIH/National Institute of Neurological Disorders and Stroke and the National Multiple Sclerosis Society. Dr. Hauser has served on the scientific advisory boards of Bionure, Annexon, Symbiotix, Neurona, and Molecular Stethoscope; has received publishing royalties from McGraw-Hill; has been an employee of the University of California, San Francisco; has received research support from NIH, the National Multiple Sclerosis Society, the Conrad Hilton Foundation; and received stock options from Bionure, Annexon, and Symbiotix (all transferred to his institution).

Dr. D'Alfonso has served on the editorial board of *Genes and Immunity* and has received research support from the Italian Ministry of Education, University and Research, the Italian Foundation for Multiple Sclerosis, and Fondazione Cariplo. Dr. Leone has served on the editorial board of the *European Journal of Neurology* and has received research support from the Italian Ministry of Health. Dr. Martinelli Boneschi has received support for congress participation or speaker honoraria from TEVA, Biogen Idec, Merck-Serono, Sanofi-Aventis and Genzyme and has received research support from Merck Serono, TEVA, Biogen Idec, Genzyme, GW-Pharma, the Italian Foundation of Multiple Sclerosis, the Italian Ministry of Health, the Fondazione Cariplo, and the European Union Seventh Framework Programme: Research and Innovation. Dr. Sorosina reports no disclosures. Dr. van der Mei has received funding for travel and speaker honoraria from Biogen Idec and has received research support from the National Health and Medical Research Council of Australia and Multiple Sclerosis Research Australia. Dr. Taylor has served on the scientific advisory boards of Biogen Idec, Australia Novartis Gilenya, and Genzyme International; has received funding for travel and speaker honoraria from Bayer Schering Pharma, MS Society of Western Australia, CSL Australia Pty. Ltd., Biogen Idec, and Novartis; has served on the editorial boards of *International MS Journal* and the *Journal of Neurology Neurosurgery and Psychiatry*; and has received research support from National Health and Medical Research Council Australia, MS Research Australia, National Health and Medical Research Council, HRC New Zealand, National MS Society of USA, the Australian Research Council, and the MS society of Tasmania. Mr. Zhou reports no disclosures. Dr. Schaefer has received research support from GlaxoSmithKline, National Institute of Environmental Health Sciences, National Institute of Mental Health, National Institute of Allergy and Infectious Diseases, National Institute on Aging, and the National Multiple Sclerosis Society. Dr. Barcellos has received research support from NIH/National Institute of Environmental Health Sciences, NIH/National Institute of Neurological Disorders and Stroke, US Environmental Protection Agency, NIH/National Cancer Institute, NIH/National Institute of Arthritis and Musculoskeletal and Skin Diseases, the National Multiple Sclerosis Society, the Peder Sather Grant for Collaborative Research, and the Robert Wood Johnson Health and Society Program Pilot Grant. Go to Neurology.org/ng for full disclosure forms.

Received December 4, 2015. Accepted in final form June 16, 2016.

REFERENCES

1. Tedeholm H, Skoog B, Lisovskaja V, Runmarker B, Nerman O, Andersen O. The outcome spectrum of multiple sclerosis: disability, mortality, and a cluster of predictors from onset. J Neurol 2015;262:1148–1163.
2. Naci H, Fleurence R, Birt J, Duhig A. Economic burden of multiple sclerosis: a systematic review of the literature. Pharmacoeconomics 2010;28:363–379.
3. Barcellos LF, Sawcer S, Ramsay PP, et al. Heterogeneity at the HLA-DRB1 locus and risk for multiple sclerosis. Hum Mol Genet 2006;15:2813–2824.
4. Patsopoulos NA, Barcellos LF, Hintzen RQ, et al. Fine-mapping the genetic association of the major histocompatibility complex in multiple sclerosis: HLA and non-HLA effects. PLoS Genet 2013;9:e1003926.
5. The International Multiple Sclerosis Genetics Consortium. Genome-wide association study of severity in multiple sclerosis. Genes Immun 2011;12:615–625.
6. Ascherio A. Environmental factors in multiple sclerosis. Expert Rev Neurother 2013;13(12 suppl):3–9.
7. Munger KL, Chitnis T, Ascherio A. Body size and risk of MS in two cohorts of US women. Neurology 2009;73:1543–1550.
8. Wingerchuk DM. Smoking: effects on multiple sclerosis susceptibility and disease progression. Ther Adv Neurol Disord 2012;5:13–22.
9. Ascherio A, Munger KL, Simon KC. Vitamin D and multiple sclerosis. Lancet Neurol 2010;9:599–612.
10. The International Multiple Sclerosis Genetics Consortium; Wellcome Trust Case Control Consortium 2, Sawcer S, et al. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. Nature 2011;476:214–219.
11. Beecham AH, Patsopoulos NA, Xifara DK, et al. Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. Nat Genet 2013;45:1353–1360.
12. Briggs FB, Acuna BS, Shen L, et al. Adverse socioeconomic position during the life course is associated with multiple sclerosis. J Epidemiol Community Health 2014;68:622–629.
13. Smestad C, Brynedal B, Jonasdottir G, et al. The impact of HLA-A and -DRB1 on age at onset, disease course and severity in Scandinavian multiple sclerosis patients. Eur J Neurol 2007;14:835–840.
14. Hedstrom AK, Baamhielm M, Olsson T, Alredsson L. Tobacco smoking, but not Swedish snuff use, increases the risk of multiple sclerosis. Neurology 2009;73:696–701.
15. Okuda DT, Srinivasan R, Oksenberg JR, et al. Genotype-phenotype correlations in multiple sclerosis: HLA genes influence disease severity inferred by 1HMR spectroscopy and MRI measures. Brain 2009;132:250–259.
16. Leone MA, Barizzone N, Esposito F, et al. Association of genetic markers with CSF oligoclonal bands in multiple sclerosis patients. PLoS One 2013;8:e64408.
17. Pittas F, Ponsonby AL, van der Mei IA, et al. Smoking is associated with progressive disease course and increased progression in clinical disability in a prospective cohort of people with multiple sclerosis. J Neurol 2009;256:577–585.
18. Simpson S Jr, Pittas F, van der Mei I, Blizzard L, Ponsonby A, Taylor B. Trends in the epidemiology of multiple sclerosis in Greater Hobart, Tasmania: 1951 to 2009. J Neurol Neurosurg Psychiatry 2011;82:180–187.
19. McDonald WI, Compston A, Edan G, et al. Recommended diagnostic criteria for multiple sclerosis: guidelines from the International Panel on the diagnosis of multiple sclerosis. Ann Neurol 2001;50:121–127.
20. De Jager PL, Chibnik LB, Cui J, et al. Integration of genetic risk factors into a clinical algorithm for multiple sclerosis susceptibility: a weighted genetic risk score. Lancet Neurol 2009;8:1111–1119.
21. Roxburgh RH, Seaman SR, Masterman T, et al. Multiple Sclerosis Severity Score: using disability and disease duration to rate disease severity. Neurology 2005;64:1144–1151.
22. Ballerini C, Guerini FR, Rombola G, et al. HLA-multiple sclerosis association in continental Italy and correlation with disease prevalence in Europe. J Neuroimmunol 2004;150:178–185.
23. Kantarci OH, de Andrade M, Weinshenker BG. Identifying disease modifying genes in multiple sclerosis. J Neuroimmunol 2002;123:144–159.
24. Tremlett H, Zhao Y, Rieckmann P, Hutchinson M. New perspectives in the natural history of multiple sclerosis. Neurology 2010;74:2004–2015.
25. Van der Walt A, Stankovich J, Bahlo M, et al. Heterogeneity at the HLA-DRB1 allelic variation locus does not influence multiple sclerosis disease severity, brain atrophy or cognition. Mult Scler 2011;17:344–352.

26. Hensiek AE, Seaman SR, Barcellos LF, et al. Familial effects on the clinical course of multiple sclerosis. Neurology 2007;68:376–383.
27. Pankratz N, Nichols WC, Elsaesser VE, et al. Alpha-synuclein and familial Parkinson's disease. Mov Disord 2009;24:1125–1131.
28. Kamboh MI, Demirci FY, Wang X, et al. Genome-wide association study of Alzheimer's disease. Transl Psychiatry 2012;2:e117.
29. Kalincik T, Guttman CRG, Krasensky J, et al. Multiple sclerosis susceptibility loci do not alter clinical and MRI outcomes in clinically isolated syndrome. Genes Immun 2013;14:244–248.
30. Hilven K, Patsopoulos NA, Dubois B, Goris A. Burden of risk variants correlates with phenotype of multiple sclerosis. Mult Scler 2015;21:1670–1680.
31. Mowry EM, Carey RF, Blasco MR, et al. Multiple sclerosis susceptibility genes: associations with relapse severity and recovery. PLoS One 2013;8:e75416.
32. Goris A, Pauwels I, Gustavsen MW, et al. Genetic variants are major determinants of CSF antibody levels in multiple sclerosis. Brain 2015;138:632–643.
33. Marta M, Stridh P, Becanovic K, et al. Multiple loci comprising immune-related genes regulate experimental neuroinflammation. Genes Immun 2010;11:21–36.
34. Beyeen AD, Adzemovic MZ, Ockinger J, et al. IL-22RA2 associates with multiple sclerosis and macrophage effector mechanisms in experimental neuroinflammation. J Immunol 2010;185:6883–6890.
35. Stridh P, Thessen Hedreul M, Beyeen AD, et al. Fine-mapping resolves Eae23 into two QTLs and implicates ZEB1 as a candidate gene regulating experimental neuroinflammation in rat. PLoS One 2010;5:e12716.
36. Adzemovic MZ, Ockinger J, Zeitelhofer M, et al. Expression of Ccl11 associates with immune response modulation and protection against neuroinflammation in rats. PLoS One 2012;7:e39794.
37. Spach KM, Case LK, Noubade R, et al. Multiple linked quantitative trait loci within the Tmevd2/Eae3 interval control the severity of experimental allergic encephalomyelitis in DBA/2J mice. Genes Immun 2010;11:649–659.
38. Weinshenker BG, Wingerchuk DM, Liu Q, et al. Genetic variation in the tumor necrosis factor alpha gene and the outcome of multiple sclerosis. Neurology 1997;49:378–385.
39. Manouchehrinia A, Helga W, Kingwell E, et al. Age related multiple sclerosis severity score: disability ranked by age. Neurology 2016;86(16 suppl):S24.007.
40. Gourraud PA, Sdika M, Khankhanian P, et al. A genome-wide association study of brain lesion distribution in multiple sclerosis. Brain 2013;136:1012–1024.